THE ROLE OF THE WSPR RESPONSE REGULATOR IN THE ADAPTIVE EVOLUTION OF EXPERIMENTAL POPULATIONS OF *PSEUDOMONAS FLUORESCENS* SBW25

Patrick Goymer

June 2002

A thesis submitted for the degree of Doctor of Philosophy

Department of Plant Sciences and University College

University of Oxford, U.K.

The Role of the WspR Response Regulator in the Adaptive Evolution of Experimental Populations of *Pseudomonas fluorescens* SBW25

Patrick Goymer, University College, Oxford
D.Phil., 2002

Abstract

The role of ecological opportunity in adaptive radiation has been demonstrated by the diversification of the bacterium *Pseudomonas fluorescens* SBW25 in a spatially structured microcosm. This provides an ideal system for studying the genetics of adaptation and asking questions about the genes that matter in evolution. Previous studies have identified the genes that are necessary for the evolved, biofilm-forming, niche-specialist genotype, the wrinkly spreader (WS). These genes are organised in two operons: the *wss* operon that encodes the genes for cellulose biosynthesis, and the *wsp* operon that encodes a chemosensory pathway. The terminal gene in the *wsp* operon, *wspR*, encodes a novel response regulator thought to regulate the activity of the *wss* operon. This gene forms the basis of this study, which assesses the role of regulatory genes in adaptive evolution. The structure-function relationship of WspR is established through the phenotypic analysis of overexpressed *wspR* random point mutants. On this basis a model of WspR activity is proposed which is tested by molecular genetic analysis. The role of phosphorylation is demonstrated by site-directed mutagenesis, and domain liberation is used to study the interaction of WspR with the other components of the signalling pathway. As the overexpression of certain *wspR* mutant alleles mimics the evolutionary transition from ancestor to niche-specialist, the fitness effects of such overexpression are measured. It is found that some, but not all, *wspR* alleles do indeed cause adaptation. It is also found that a phenotypically-plastic genotype, with enhanced fitness, can be created by artificial manipulation, but does not occur naturally; this demonstrates the existence of a constraint on evolution. Sequence analysis of independently-isolated WS genotypes shows no evidence of *wspR* sequence variation, despite its capacity to enhance fitness. A further proteomic and phenotypic characterisation shows variation between ancestral and WS genotypes, and also between different WS genotypes. This demonstrates that there are different mutational routes to the same adaptation.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr Paul Rainey, for giving me the opportunity to work on such an exciting project, and for his moral and intellectual support. I would also like to thank my second supervisor, Dr Malcolm Campbell, for much advice. The BBSRC and University College have provided me with the financial assistance that made the work possible.

I would like to thank all the members of the Rainey lab, past and present, for their assistance. I am particularly grateful to Jane Bond, Julie Stansfield and Zena Robinson for technical support; to Dr Andrew Spiers and Dr Angus Buckling for constant advice; and to Dr Robert Jackson and Dr Rees Kassen for very helpful comments on the thesis.

Finally I would like to thank all the friends and family who have made my the last three years as enjoyable as they have been. I am particularly grateful to those friends who have been in Oxford with me during that time: first and foremost, Helen, and also Graham, Andrew, Maggie, Mark, Andy, Julian, Helena, Frank, Anna, Steve, Edd, and many others.

Table of Contents

Table of Abbreviations

ANOVA	Analysis of Variance
Ap	Ampicillin
c-di-GMP	Cyclic-bis(3'→5') diguanylic acid
CLP	Cellulose-Like Polymer
CR	Congo Red
CV	Crystal Violet
EPS	Extracellular Polysaccharide
FS	Fuzzy Spreader
HK	Histidine Kinase
IPTG	Isopropyl-β-D-galactopyranoside
KB	King's B
Km	Kanamycin
LB	Luria-Bertani
LPS	Lipopolysaccharide
LSWS	Large Spreading Wrinkly Spreader
MALDI-TOF	Matrix-assisted Laser Desorption Ionisation Time-of-flight
MCP	Methyl-accepting Chemotaxis Protein
OD	Optical Density
ONPG	o-Nitrophenyl-β-D-galactopyranoside
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
Q-TOF	Quadropole Time-of-flight
RR	Response Regulator
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SM	Smooth
SOE	Strand Overlap Extension
Tc	Tetracycline
WS	Wrinkly Spreader
wsp	wrinkly spreader phenotype
wss	wrinkly spreader structural
X-Gal	5-Bromo-4-chloro-3-indolyl- β-D-galactopyranoside

Chapter 1

Introduction

1.1. Overview of the genetics of adaptation

The genetic basis of adaptation has been the subject of study ever since the modern synthesis (Fisher, 1930; reviewed by Provine, 1971) that wedded Darwin's theory of natural selection (Darwin, 1859) to the genetics of Mendel, as promoted by Bateson (1909) and others. This interest therefore precedes the discovery of the physical nature of the gene (Avery *et al.*, 1944; Watson and Crick, 1953) and its function in encoding proteins (Crick *et al.*,1961, Beadle and Tatum, 1941). Despite such a long period of study, encompassing enormously fruitful discoveries in related fields, little progress has been made over this time in identifying the genes and mutations that are responsible for adaptive evolution.

There are several major problems that are in part responsible for this lack of progress, some of which have diminished significantly with recent advances. Most of the problems involve the difficulty of extracting relevant data from vast amounts of information, much of it incomplete. The study of evolution has normally relied, on the one hand, on the use of fossils, for which adaptations can be identified but there is no genetic information, and, on the other hand, comparative studies of extant species, where separating adaptive differences at the level of the DNA sequence from neutral ones is like finding the proverbial needle in the haystack. Even in the latter case it is only in very recent times that mass sequencing and genome projects have allowed the full range of genetic differences to be studied. On top of this it has tended to be the policy to sequence only genomes of species that are not closely related, in order to achieve a broad range of mechanistic analyses. The assumption that closely related organisms will have similar biology

may or may not be correct but, even where it is, this policy is unhelpful to the evolutionary biologist, for whom closely related organisms present the best opportunity for progress; this is because more is known about the evolutionary and environmental history separating closely related species so that adaptive differences are more easily disentangled from neutral differences.

The possibility of a genetic theory of adaptation is most likely to come from the experimental evolutionary approach. The tools of molecular biology, sequencing and functional genomics, when coupled with the classical controlled evolutionary manipulations pioneered by Lack (1947) and the use of prokaryotic models of evolution (e.g. Lenski *et al.*, 1991; Rainey and Travisano, 1998), will be invaluable in showing us what types of mutation are selected to cause adaptation and the factors affecting them. The search for such a theory is discussed at greater length by Travisano (2001), especially in reference to prokaryotic models. This is highly relevant to this study which is an examination of the genetics behind a particular prokaryotic adaptation isolated in a controlled experiment.

1.2. Microevolutionary studies of adaptive genetics

Fisher's model of evolutionary genetics (Fisher, 1930) argues that adaptation occurs through the fixation of many genes of small effect. The rationale behind this is that mutations of large effect are far more likely to be deleterious, especially when the multidimensional complexity of an organism is considered. This argument is shown in diagrammatic form in Figure 1.1. However, more recent authors have questioned these assumptions on both theoretical and experimental grounds. Kimura (1983) points out that the advantage of small-effect alleles due to their increased chance of being beneficial may be outweighed by their decreased chance of being fixed in the population. Stern (2000) questions the assumption of universal pleiotropy in Fisher's model, pointing out that if genes have an effect on one character alone then the size of the mutation is not

important, assuming they are not too close to a fitness maximum (see Figure 1.1). Orr (1999) provides a simulated model of adaptation showing exponential distributions of gene effects, which perhaps argues for a more balanced view.

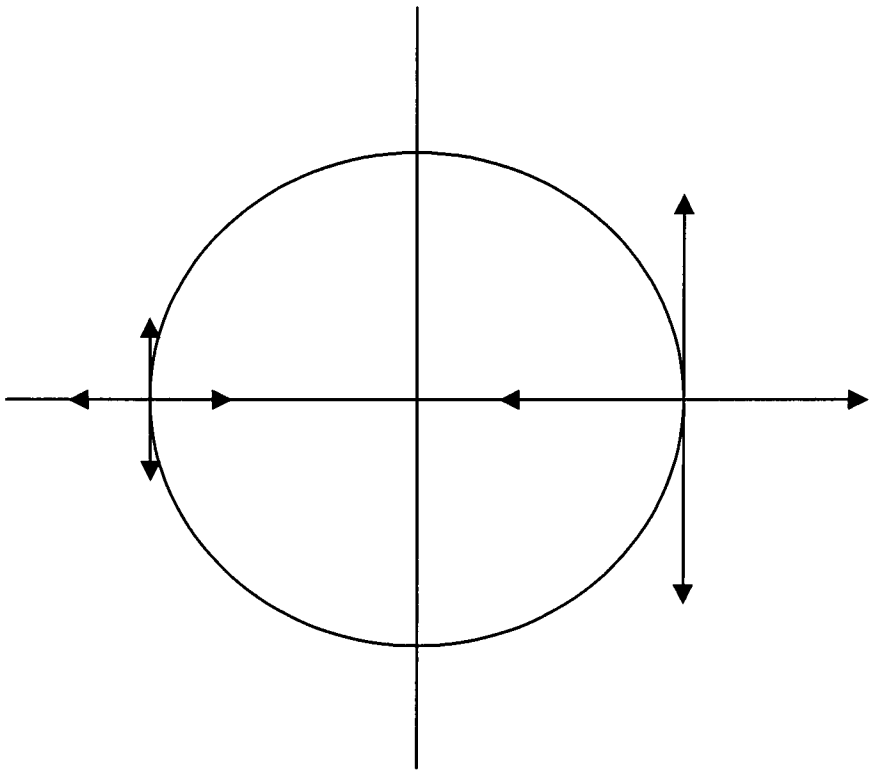


Figure 1.1. Diagrammatic representation of Fisher's micromutationist argument. In this diagram Fisher's argument that a mutation affecting one character will also affect others is reduced to the case where there are only two characters. The axes represent these two characters with the origin being the point of optimal fitness. On the right hand side of the diagram the organism is supposed to be located where the circle crosses the x axis. A large mutation occurs in a random direction and the arrows represent the fact that the probability of the organism becoming closer to the origin (i.e. moving inside the circle) is less than half. The left hand side of the diagram shows an equivalent situation with a much smaller size mutation. The probability of moving inside the circle in this case is significantly greater and in fact increases to a half as the size of the mutation decreases to zero. This argument becomes stronger as more dimensions are added. This represents the pleiotropic nature of the mutations. However, if the mutation affects just one trait there is only one dimension, represented by the x axis in the diagram. In this case, with the exception of mutations that are so large that they overshoot the origin, all mutations have a one in two chance of increasing fitness.

Orr and Coyne (1992) assess the evidence for macro- as opposed to micromutationism. They conclude that some, though by no means all, adaptations arise partly as a result of large mutations. These mutations do not, however, produce the adaptation *de novo* as in the theory of Goldschmidt (1940), but cover a substantial amount of the phenotypic space between ancestral

and evolved forms. The sources of evidence that they consider are the few experimental studies that have revealed the genetic basis of adaptations. Some studies have confirmed the role of large mutations in cases such as mimicry (Charlesworth, 1994), melanism (Turner, 1977), parasite resistance in *Drosophila melanogaster* (Orr and Irving, 1997) and heat tolerance in bacteriophage φX174 (Bull *et al.*, 2000). On the other hand some studies have shown the effect of small mutations, for example Burch and Chao (1999) in an RNA phage, and Levin *et al.* (2000) in *Escherichia coli*. Hence it remains an unresolved question, and probably one with a complicated answer, which the unravelling of an adaptive event at a mechanistic level may help to resolve. Such an unravelling, coupled with experimental manipulation, may help to answer the related question of what genetic and environmental factors affect the course of adaptive evolution in a population.

The last three studies cited above are pertinent because they refer to prokaryotic systems. The present study uses a bacterial model and, while it is hoped that much will be of general relevance, it should be noted as a caveat that bacteria differ from eukaryotes in many respects. The use of prokaryotic models will be discussed in greater detail below, but some of their genetic peculiarities deserve mention here. Those that are evolutionary relevant are usefully reviewed in Levin and Bergstrom (2000). The two major differences, when contrasted with sexual eukaryotes, are the low levels of recombination in bacteria and the high levels and phylogenetic breadth of horizontal gene transfer. The differences caused by haploidy, as opposed to the diploidy that is the norm in higher eukaryotes, are illustrated by Paquin and Adams (1983). They show that the rate of evolution is almost doubled in equivalent diploid populations. Because of the lack of vertical sex in bacteria, and the resulting difficulty in defining species, macro- and microevolution are not as distinct as they are in eukaryotes.

1.3. Development and macroevolutionary adaptive genetics

The introduction so far has concentrated on adaptation at a microevolutionary level. While this is assumed to form the basis for all evolutionary advances, it is of interest to look at comparisons between species when studying adaptation, as the differences may be larger and more obvious. Though this study concerns bacteria, some of the best examples of adaptive differences between species come form multicellular organisms where the genetics are those of developmental biology. This is because developmental biology is the process by which genotype is translated into phenotype. The desire to look at both micro- and macroevolutionary scales will become clear when the experimental system is described below; though the timescale studied is short it is an adaptive radiation that is being considered, as in asexual organisms different genotypes are effectively different species. This is the stuff of macroevolution, though, as mentioned above, the distinction between the two processes is less clear-cut in an asexual organism.

The purpose of this study is to elucidate the links between genotype and phenotype and between phenotype and fitness, all in the same system. The first of these links is the study of developmental biology and recently this has been looked at in the context of the second link (reviewed in Purugganan, 1998). Stern (2000) considers which genes are responsible for adaptive differences in body plans. He concludes that the *cis*-elements of regulatory genes, which have a highly modular form, may play a large role, and escape the problem of pleiotropy and its requirement for micromutationism. This approach mirrors that of Golding and Dean (1998) that looks at similar modularity in protein structure and its evolutionary consequences.

Akam (1998), however, is more micromutationist in his assessment of development by proposing a model of homeobox domain gene evolution that is more subtle than the homeotic mutations observed in the laboratory. The diversity of regulatory elements in *hox* clusters allows for gradual

evolution of segment identity without the need for 'hopeless monsters'. The key point in this study is that mutations that are engineered in the laboratory are not necessarily the ones that have occurred during evolution. This problem is greater in organisms where fitness is harder to measure than it is in the prokaryotic system used in this study, in which easy fitness measurements can distinguish evolutionary plausible mutations from the so-called 'hopeless monsters'. This will be discussed further in Chapter 5.

Tautz and Schmid (1998) point out that the genes most often studied by developmental biologists are those that affect early embryos and are therefore highly conserved across phyla. These are unlikely to be the genes that affect short-term ecological adaptations as the latter are more likely to be modifications to adult morphology. These authors suggest that the major macroevolutionary changes may result from mutations in the embryonic genes, while microevolutionary adaptations result from mutations in adult genes. The terminology is questionable as the latter are likely to be involved in all but the most dramatic speciation events but the key point is the suggestion that regulatory genes are responsible for much adaptation. This is then elaborated further by Tautz (2000) and Purugganan (2000), both of whom suggest that regulatory gene evolution is the driving force behind adaptation. This is as likely to hold for prokaryotes as it is for eukaryotes, and probably for both micro- and macroevolution. Also Kirschner and Gerhart (1998) point out that the evolvability of an organism depends on the molecular mechanisms behind its development. They too implicate regulatory genes in the generation of phenotypic diversity as they are less constrained.

The experimental studies that have been carried out implicating regulatory genes in adaptive evolution include the following. Dickinson *et al.* (1984) show that there are adaptive differences in expression levels of regulatory genes in closely related *Drosophila* species, while Palopoli and Patel (1998) look at adaptive differences in *hox* genes in a similar range of dipterans. Belting *et*

al. (1998) provide evidence that *hox* gene *cis*-elements are also involved in adaptive differences between mice and chickens, a far larger phylogenetic distance, encompassing far more radical body plan differences than those within the diptera. Purugganan and Suddith (1998) show a higher than expected amount of non-neutral variation in the *Arabadopsis thaliana* CAULIFLOWER regulatory gene in comparison with structural and housekeeping genes. Most notably of all, Barrier *et al.* (2001) show that the homologues of the *Arabidopsis* APETALA gene show accelerated evolution over structural genes in the adaptive radiation of the species in the Hawaiian silversword alliance. In these studies it is important to distinguish between evolution in coding and non-coding regions. Doebley and Lukens (1998) propose that the latter play a more important role as they are more evolvable, and Carroll (2000) holds the same view. Neither author, however, rules out mutations in coding regions, especially at points in pathways where constraint is light.

Purugganan (2000) poses some questions that need to be answered about the evolution of regulatory genes, which include how much constraint acts at different points in a pathway, and whether any differences are systematic across pathways. It may be that adaptive mutations are more likely to appear at the beginning of regulatory pathways rather than at the end, or that the earlier large-effect mutations occur there and the later fine-tuning occurs further down the pathway. Alternatively there may be few general rules due to the great diversity of potential cross-talk in different pathways. These are the kinds of questions that make the evolutionary genetics of such loci interesting.

1.4. Development applied to prokaryotes

The above consideration of developmental biology may be considered irrelevant to a study of bacteria because they are unicellular organisms with no body plan. However, if evolution often

proceeds by the selection of mutations at regulatory loci, bacteria are just as representative as any particular 'higher' eukaryote. While bacteria lack significant spatial patterning of gene expression of the kind that allows *hox* genes to pattern animal segments, they do regulate their gene expression temporally (and in fact show some intracellular spatial patterning). Indeed it can be argued that development consists solely of a series of cell-fate decisions, and that the best studied such decision is that of bacteriophage-λ (Ptashne, 1992). In this case the precise molecular explanation of the decision between lytic and lysogenic life cycles is understood. As such it can be considered a paradigm for developmental genetics. Another pertinent example is that of *Caulobacter crescentus*, which has two cell types in its life cycle. The switch between the two is governed in part by the response regulator PleD (Aldridge and Jenal, 1999), a homologue of WspR, the main focus of this study. For this reason it will be discussed in greater detail below.

The unicellular nature of the bacterial life cycle, and the subsequent lack of spatial patterning, is in fact a simplification. The system used in this study involves the formation of a bacterial biofilm. Several authors (e.g. O'Toole *et al.*, 2000) have postulated biofilms as primitive forms of multicellularity, the presence and role of which in the prokaryotic world has been studied by Shapiro (1998). A further significant advantage to studying bacteria is the ease with which fitness measurements can be made. When developmental mutations are produced in experimental metazoan organisms a series of life-history trait measurements have to be made to estimate fitness in a natural environment, and hence assess the possibility of evolution following that route. In the system to be described below direct measurements of fitness can be made, in competition with the ancestor, in the environment in which the evolutionary event of interest originally occurred.

Finally, the work of Pao and Saier (1995) looks at the evolution of bacterial response regulator proteins. As will be discussed below, the main focus of this study is the response regulator WspR. These authors show evidence for domain shuffling during the evolution of these proteins, a

process thought to have occurred in the evolution of *wspR*. It therefore looks probable that such proteins are an important source of adaptation in bacteria, as they are thought to be in eukaryotes.

1.5 The adaptive radiation of *Pseudomonas fluorescens*

It has been noted that the *Pseudomonas* genus shows arguably the most diversity in the bacterial kingdom (Spiers *et al.*, 2000). These authors go on to argue that there is no special reason for this, and it can be attributed to a long evolutionary history and a wide range of ecological opportunity. These factors combine to make a species from this genus a good model in which to study adaptive radiation, because of its diverse range of characterised relatives and its representativeness of other examples.

The system that forms the basis of this study is described by Rainey and Travisano (1998). An overview of the relevant factors shall be given here. The experiment followed a single genotype as it evolved to fill both a spatially structured environment, and one that was unstructured. The rationale behind this comparison was to test the importance of ecological opportunity in diversification. Within the structured environment there are a range of different ecological niches to which it was hoped to observe adaptation. In such a controlled environment the role of competition in promoting and maintaining diversity could be assessed.

The bacterium used was *Pseudomonas fluorescens* SBW25 (Rainey and Bailey, 1996), a common aerobic plant coloniser. The structured environment in question was a static glass microcosm containing 6ml of the nutrient-rich broth, King's medium B. This environment was inoculated with 6μl of ancestral bacteria, which has a smooth colony morphology when grown on agar plates. Over a period of 10 days, using replicate microcosms, the amount of diversity in colony morphology seen when diluted samples were plated out was measured using the Shannon-Weaver

index (Shannon and Weaver,1949). This diversity, and that in the control samples from spatially unstructured (shaken) microcosms, is shown in Figure 1.2. Also shown are the relative abundance of the three principle morphs. It can be seen that a structured environment promotes the increase of diversity over time and that by the end of the experimental time period the most abundant morph has a different phenotype from the ancestor. None of this occurs in the spatially unstructured environment.

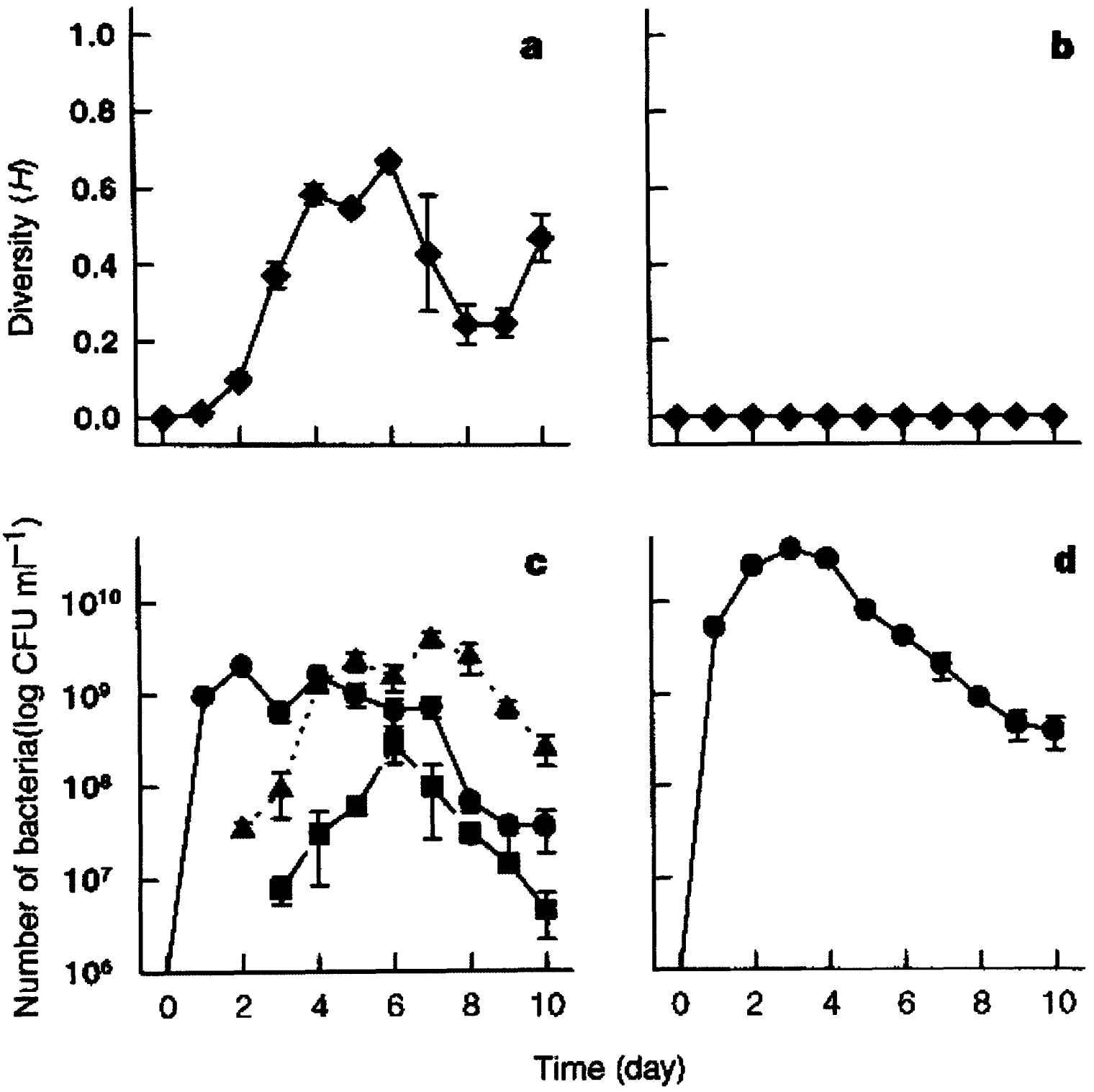


Figure 1.2. Effect of ecological opportunity on the evolution of genetic diversity. Two sets of identical populations were founded from the ancestral smooth morph and incubated either with or without shaking. Every 24 h, three replicate microcosms were destructively collected and diversity was determined by scoring the frequencies of morphs. Panels a and b: Diversity increased rapidly in the structured environment (a) whereas no diversity was detected in environments lacking spatial structure (b). Panels c and d: Evolutionary dynamics of the principal morph classes in the structured (c) and unstructured (d) environments. Values are the means +/-s.e.m (n = 3). Circles represent smooth morphs, triangles represent wrinkly-spreader morphs and squares represent fuzzy-spreader morphs. Figure and legend reproduced from Rainey and Travisano (1998).

It was found that most of the colony morphologies observed fell into three classes: the wrinkly spreader (WS; the most abundant morph), the smooth (SM; indistinguishable from the ancestor), and the fuzzy spreader (FS). Moreover it was discovered that there was a fortuitous correlation between colony morphology on agar and niche inhabited in the microcosm. This is illustrated in Figure 1.3. The wrinkly spreader forms a mat at the air-broth interface, the fuzzy spreader colonises the bottom of the microcosm, and the smooth occupies the main bulk of the liquid. Within each of these classes there are multiple genotypes.

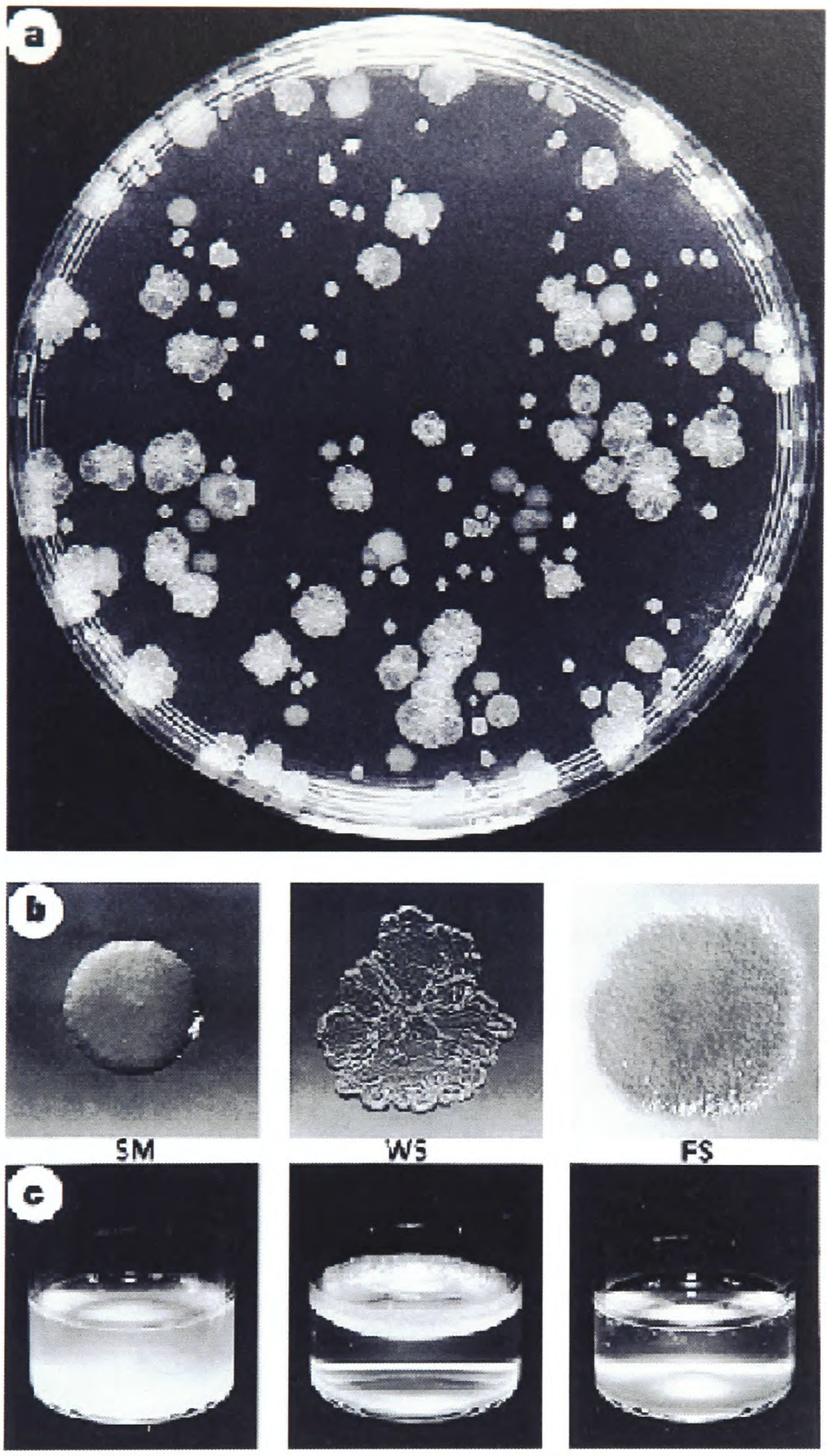


Figure 1.3. Phenotypic diversity and niche specificity among *P. fluorescens* SBW25 colonies evolved in a spatially structured environment. Populations were founded from single ancestral 'smooth' (SM morph) cells and propagated in 6-ml King's medium B contained in a 25-ml microcosm at 28 °C. Microcosms were incubated without shaking to produce a spatially structured environment. Panel a: After 7 days, populations show substantial phenotypic diversity which is seen after plating. Panel b: Most phenotypic variants can be assigned to one of three principle morph classes: (SM), wrinkly spreader (WS) and fuzzy spreader (FS). Panel c: Evolved morphs showed marked niche preferences. Figure and legend reproduced from Rainey and Travisano (1998).

These niche specialisations were found to be heritable phenotypes, maintained in the population by negative frequency-dependent selection. This indicates that the driving force behind adaptive radiation was competition. These factors make this system representative enough to be worthy of genetic study, and at the same time sufficiently simple to be tractable. Indeed it has already been used for studies showing that nutrient abundance affects diversification (Travisano and Rainey, 2000), and that diversity peaks at intermediate levels of disturbance (Buckling *et al.*, 2000) and productivity (Kassen *et al.*, 2000).

1.6 Further studies of adaptive radiation and niche colonisation

The issues raised in the studies discussed above are those of ecological specialisation and competition. These issues have been considered by other authors both theoretically (Abrams, 1987; Futuyma and Moreno, 1988) and experimentally (Cooper and Lenski, 2000). The majority of the work in the latter case has been in experimental prokaryotic systems, the advantages of which are discussed by Rainey *et al.* (2000) and Travisano (2001). An exception to this rule is a study showing stickleback divergence promoted by competition (Schluter, 1994). Two of the prokaryotic studies show ecological specialisation evolving not due to a spatially structured environment, but due to frequency-dependent metabolic interactions between genotypes (Turner *et al.*, 1996; Rosenzweig *et al.*, 1994). However, though these studies encompass a wide range of examples that confirm the general relevance of the issues, none of them provide a genetic explanation for the initial radiation nor for the competitive trade-offs that maintain it. Therefore the next consideration shall be the genetics of the *P. fluorescens* system and, in Chapter 5, fitness trade-offs and the evolution of specialists and generalists shall be considered in more detail.

1.7. The genetics of the wrinkly spreader

As the most abundant, and hence possibly the fittest, morph the wrinkly spreader (WS) was selected for a genetic analysis (Kahn, 1998). As there was diversity in colony morphology and fitness within this class an isolate that was deemed most representative was selected. It is quite possible that there are several genetic routes to similar niche-specialising phenotypes, reflected in the range of morphs, but for the meantime at least it was necessary to select just one. This issue will be returned to in Chapter 7. The selected genotype was named the large spreading wrinkly spreader (LSWS) or, more formally, PR1200 (Spiers *et al.*, 2002).

Transposon mutagenesis was carried out on both ancestral SM and LSWS genotypes, looking for WS morphs in the former case and SM revertants in the latter (Kahn, 1998). This kind of mutagenesis, here performed using a mini-Tn5-Km transposon (Simon *et al.*, 1983; Herrero *et al.*, 1990), inactivates genes resulting in loss-of-function phenotypes, unlike chemical mutagenesis that may alter single nucleotides. When the two mutageneses were performed only that of the LSWS produced the expected mutant colonies. This strongly implies that the mutation that caused the initial SM-WS transition was gain-of-function. Had it been the case that only the mutagenesis of the ancestor produced results the mutation that produced the wrinkly spreader would have been assumed to have been loss-of-function. Had both mutageneses produced results, one of the two would have been due to hits in modifying genes and no such inference could have been drawn. It should also be noted that these transposons contain transcriptional terminators, thus they affect genes downstream in the operon as well. Because of this polar effect it is not always the gene with the insertion whose disruption causes the mutant phenotype.

The screen produced 8000 mutants that were kanamycin resistant, of which 40 had reverted to the ancestral colony morphology. This number was reduced to 27 when siblings had been removed

and more extensive phenotypic checks had been performed. The crucial phenotypic test was for loss of niche-preference as this is the trait of evolutionary importance. These 27 mutants were then classified on phenotypic and genetic grounds into 6 groups (Spiers *et al.*, 2002). The phenotypic characterisation involved firstly studies of colony morphology, and then studies of diversification. In the latter case the mutants were used to inoculate microcosms (containing kanamycin to maintain the transposon) as in the original experiments of Rainey and Travisano (1998). The amount of diversity produced was compared with that resulting from a wild-type ancestor. One group of mutants diversified fully like the ancestor; another diversified but failed to produce the WS phenotype; the third type failed to diversify at all. This last group was divided into three on the basis of the colony morphology observations. The genetic characterisation involved the physical mapping of the mutants using an existing map (Rainey and Bailey, 1996). The six groups arising from this classification corresponded closely with those from the phenotypic analysis. The groups and the genes in them are shown in Table 1.1.

Group	Number of mutations in group	Colony Morphology	Diversification	Biofilm formation	Identity of characterised mutants	Name of mutated gene	Proposed function	Cellulose production
I	6	SM	S+W+F	-	WS4	*wspR*	Response regulator	-
II	14	SM	S+F	-	WS1	*wssA*	Spatial localisation	-
					WS13	*wssB*	Cellulose synthase subunit	-
					WS22	*wssC*	Cellulose synthase subunit	-
					WS15	*wssD*	Endoglucanase (cellulase)	-
					WS25	*wssE*	Cellulose synthase subunit	-
III	3	SM*	S	-	WS18	*wssF*	Unknown function	+
					WS6	*wssH*	Polymer modification	+
					WS9	*wssH*	Polymer modification	+
IV	2	SM*	S	-	WS12	*pgiA*	Phosphoglucose isomerase	+
V	1	SM	S	+	WS39	*mreB*	Murein biosynthesis	+
VI	1	SM*	S	+	WS5	ND	ND	+

Table 1.1. Characterisation of WS mutants. The 27 mutations that caused the LSWS to revert to the ancestral state are shown, classified into their 6 groups. They were isolated for their smooth colony morphology, but those marked SM* showed wrinkly characteristics after several days growth on plates. When evolved in static microcosms the different mutants showed different levels of diversity. The presence of smooths, wrinklies or fuzzies is indicated by S, W and F. Classes V and VI showed weak biofilm formation (as assessed qualitatively by eye) unlike the ancestor, but less strong than the LSWS morph. Certain mutants were characterised physically and their names, the genes they disrupt and the functional assignment of those genes are shown. No gene assignment has been made for WS5, hence it is listed ND (not determined). Also shown is the ability to produce cellulose assessed by Congo Red staining.

The WS-4 mutation in class I is *wspR*, the focus of this study. It is notable that a mutation at this locus does not prevent diversification, implying the role of a regulatory gene that can be bypassed by mutations at other regulatory loci. This locus will be discussed in greater detail below.

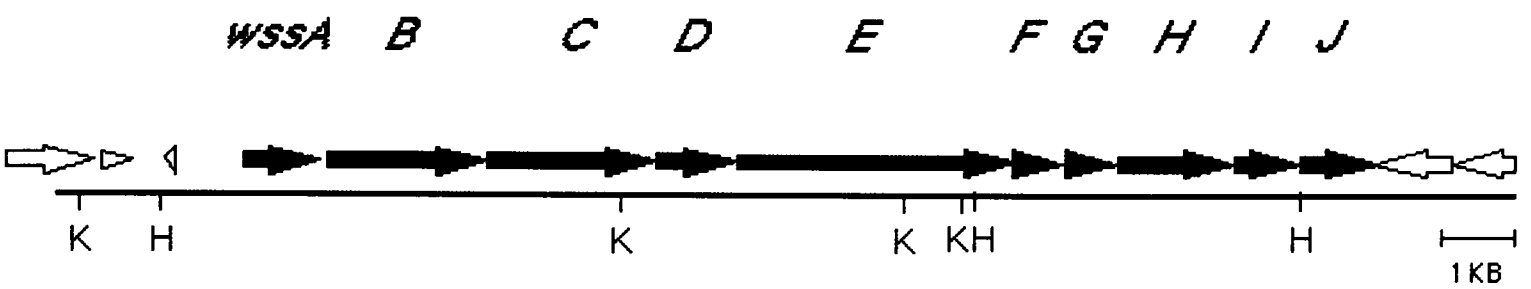


Figure 1.4. Genetic structure of the wss operon. The *wss* operon is a cluster of ten genes predicted to encode the proteins that synthesise a cellulose-like polymer (CLP). The cellulose synthase is predicted to be comprised of WssB, WssC and WssE; WssD is an associated cellulase; WssG, WssH and WssI are predicted to play a role in acetylation of the basic cellulose polymer. WssA and WssJ have homology with MinD. Upstream of the *wss* operon is tRNAThr (open arrow), a truncated gene ('fabG) encoding an oxidoreductase and a gene predicted to encode a membrane protein porin Q (oprQ). Downstream of the operon is a putative arginine and orinithine transport gene (*aotM*) and a hypothetical protein of no known function, PA5156. The locations of certain restriction sites are indicated: H, *Hin*dIII; K, *Kpn*I. The position of the *wssE*-'*lacZ* fusion is at the *Hin*dIII site at the end of *wssE*; Figure and legend obtained from A. Spiers (personal communication).

Groups II and III contain the genes that together constitute the *wss* operon (Figure 1.4). Mutations in the former do not prevent diversification *per se,* but do prevent the formation of the WS phenotype. On these grounds the operon was first assumed (Kahn, 1998) to encode the structural genes that encode the selected phenotype. Further analysis (Spiers *et al.*, 2002) has confirmed this and *in silico* analysis has identified the genes in group II as homologues of cellulose synthesis genes in other bacteria. The affinity of WS genotypes for Congo Red and calcofluor, known to bind the β-1,4 glucan bond of cellulose, has lead to the extracellular component of the WS mat being referred to as cellulose-like polymer (CLP), a fact confirmed by spectroscopic analysis (A. Spiers, personal communication). The group III mutations at the end of the operon are thought to be in genes that modify the CLP (Spiers *et al.*, 2002). They have homology to alginate synthesis genes, which have been implicated in the production of biofilms in *Pseudomonas aeruginosa* (Davies and Geesey, 1995).

1.8. Bacterial cellulose biosynthesis

Cellulose is one of the most abundant natural polymers, being produced by most plants. It is also produced by a number of bacterial species. Fortunately for this study, these systems have been studied as paradigms for the situation in plants because of their relative simplicity and amenability to laboratory investigation. The most studied example is that of *Acetobacter xylinum* (Swissa *et al.*, 1980). These authors describe the metabolic pathway involved in cellulose synthesis: Glucose is taken through the membrane and converted to glucose-6-phosphate by glucose kinase; this is isomerised to glucose-1-phosphate by phosphoglucomutase and then converted to uridine 5'-diphosphate glucose (UDPG) by UDPG pyrophosphorylase; the last and most critical step is the polymerisation of UDPG into cellulose by cellulose synthase. It is this step that is the only unique one in the pathway and is therefore the site of regulation. Wong *et al.* (1990) and Saxena *et al.* (1994) describe the operons that encode cellulose synthase in two different strains of *A. xylinum*. It is the former of these two that is homologous to the *wss* operon. The genes *bcsA*, *bcsB*, and *bcsC*, which correspond to *wssB*, *wssC*, and *wssE* respectively, form the cellulose synthase complex, with BcsB being the cyclic-di-GMP binding protein to be discussed more fully below. The cellulose synthase complex forms a basket-like pore in the bacterial membrane through which ribbon-like microfibrils are synthesised out of the cell (Brown, 1996). Interestingly, the ratio of the predominant allomorph, cellulose I, to its less common counterpart, cellulose II, can be altered by mutations in the *acsD* gene (Saxena *et al.*, 1994); such modifications to the structure of the extracellular matrix, like those performed by the products of the genes *wssF* to *wssH*, are possible targets for selection.

Aside from *A. xylinum* the genetics of cellulose synthesis have been studied in several other bacterial species. In *Agrobacterium tumefaciens* the two *cel* operons (Matthysse *et al.*, 1995a) encode the enzymes involved in cellulose polymerisation (Matthysse *et al.*, 1995b). In this plant

pathogen the cellulose microfibrils are involved in attachment to the plant during infection (Matthysse, 1983). There is some evidence (Massey, Gal, Preston and Rainey, unpublished data) that this may be the role of cellulose in natural populations of *P. fluorescens*, though it is not the use that is made of it in microcosms.

The structural genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii* show both sequence and organisational homology to those in *A. tumefaciens* (Ausmees *et al.*, 1999). In addition, these authors describe a pair of regulatory genes at a tandemly located locus, *celR1* and *celR2*. These have sequence similarity to *divK* and *pleD* of *C. crescentus* respectively, and by extension *celR2* is homologous to *wspR*. In the enteric bacteria Zogaj *et al.* (2001) have shown that the multicellular morphotypes of *Salmonella* species and *E. coli* produce cellulose in their extracellular matrices. The production in *S. typhimurium* is regulated by AdrA, the product of another gene with similarity to *wspR*. These are the first cases in which the regulatory elements for cellulose synthesis have been discovered and will be discussed more fully below with respect to *wspR*.

An interesting aside on cellulose production in unicellular organisms is the observation by Grimson *et al.* (1996) that cellulose plays a role in the development of the eukaryotic slime mold *Dictyostelium discoideum*. This organism bridges the gap between unicellular and multicellular modes of life and is therefore interesting in the light of the comments on biofilms and multicellularity made above.

1.9 Biofilms

In the WS phenotype the fitness-enhancing factor is the ability to form a biofilm at the air-broth interface. This particular biofilm requires the production of CLP, whose synthesis is described

above. It is of note that the cellulose synthase genes are probably present in *P. fluorescens* for plant colonisation, but that in this novel environment they have been exapted for biofilm formation, more akin to the enteric bacteria. Here I shall consider some general properties of biofilms, regardless of the physical nature of the attachment matrix.

In recent years the importance of biofilms in microbial ecology has become more and more apparent (Davey and O'Toole, 2000). It seems likely that the planktonic paradigm of bacterial existence is only one part of the truth and that complex communities are more typical. Examples of biofilms in nature are those formed in mammalian infections (Dickinson and Bisno, 1993), those formed in the plant rhizosphere (Davey and O'Toole, 2000), and those found in aqueous (Overmann and Pfennig, 1992) and extreme (Edwards *et al.*, 2000) environments. Many natural biofilms are multi-species structures with complex co-operative and competitive interactions, though some single-species examples are found (O'Toole *et al.*, 2000). Single-species biofilms have been the main focus of genetic study due to the complexity of the interactions involved in the formation of multi-species cases. However, there is evidence that the structural basis is similar in the two cases (Costerton *et al.*, 1995), even if the genetics are more complicated in the latter.

Biofilms are not simply homogenous aggregates of bacteria bound by an extracellular matrix, they possess characteristic structural features. O'Toole *et al.* (2000) review the formation of a *P. aeruginosa* biofilm, this species being the most studied structurally and genetically. *P. aeruginosa* cells find the surface on which the biofilm will form by flagellar swimming motility but then move across the surface attaching to other cells using the type IV pili-dependent twitching motility (O'Toole and Kolter, 1998a). Attachment induces the expression of alginate biosynthesis genes that produce large amounts of this extracellular polysaccharide (EPS). Maturation of the biofilm involves the formation of mushroom-shaped microcolonies, linked together by the EPS. This process is dependent on *N*-acyl-homoserine lactones, implicating

quorum sensing in the architecture of the biofilm, though in *P. fluorescens* longer fatty acyl chains are thought to play a similar role (Allison *et al.*, 1998). A model of biofilm formation in various organisms is shown in Figure 1.5.

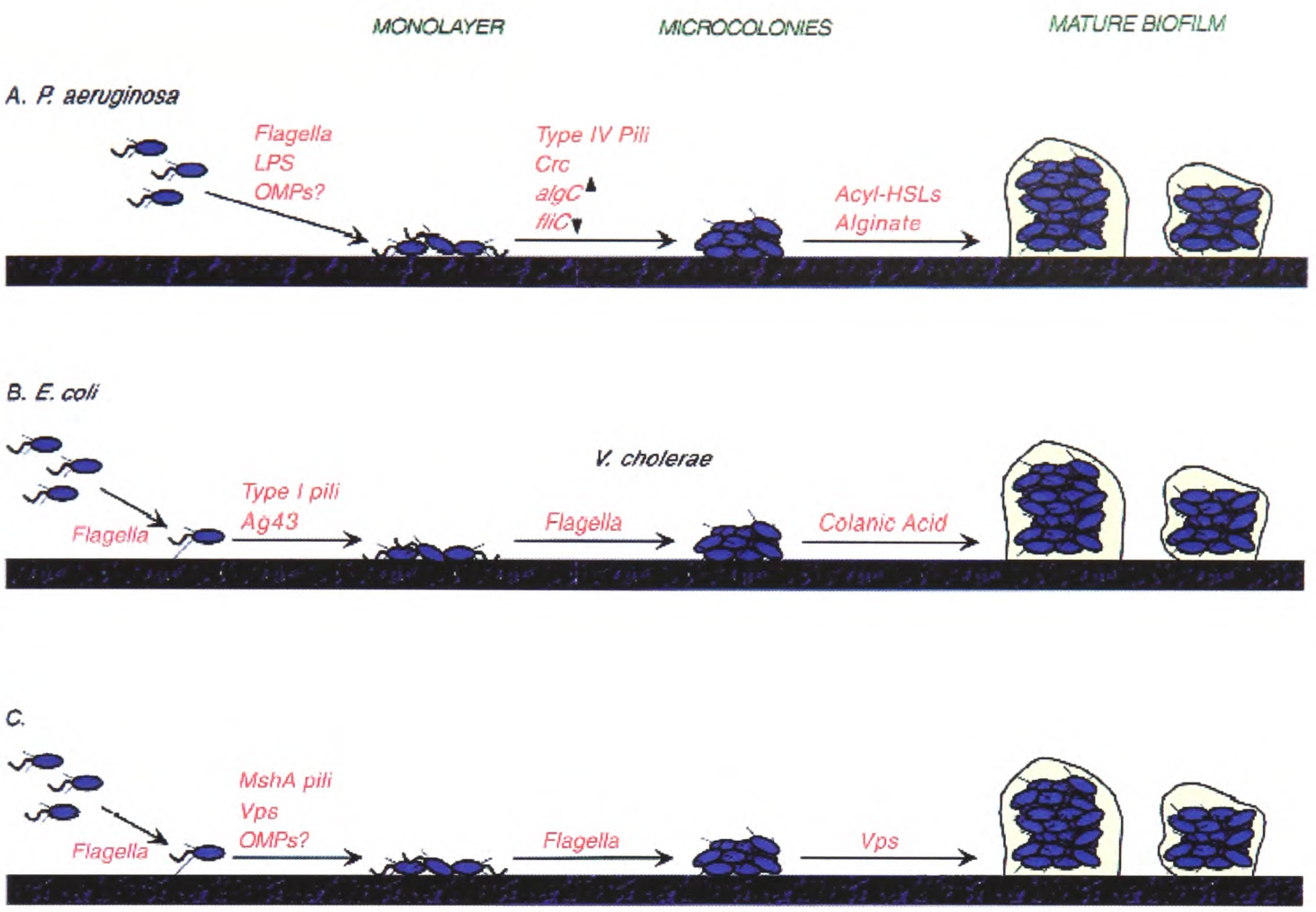


Figure 1.5. Biofilm development in gram-negative organisms. This figure outlines the current models for the early stages in biofilm formation in three of the best-studied model organisms, *P. aeruginosa, E. coli*, and *V. cholerae*. (A) In *P. aeruginosa*, flagella are required to bring the bacterium into proximity with the surface, and LPS mediates early interactions, with an additional possible role for outer membrane proteins (OMPs). Once bacteria are on the surface in a monolayer, type IV pilus-mediated twitching motility is required for the cells to aggregate into microcolonies. The production of pili is regulated at least in part by nutritional signals via Crc. Documented changes in gene expression at this early stage include up-regulation of the alginate biosynthesis genes and down-regulation of flagellar synthesis. The production of cell-to-cell signalling molecules (acyl-HSLs) is required for formation of the mature biofilm. Alginate may also play a structural role in this process. (B.) In *E. coli*, flagellum-mediated swimming is required for both approaching and moving across the surface. Organism-surface interactions require type I pili and the outer membrane protein Ag43. Finally, the EPS known as colanic acid is required for development of the normal *E. coli* biofilm architecture. (C) *V. cholerae*, like *E. coli*, utilises the flagella to approach and spread across the surface. The MshA pili, and possibly one or more unidentified outer membrane proteins, are required for attachment to the surface. This initial surface attachment appears to be stabilised by EPS. Formation of the mature biofilm, with its associated three-dimensional structure, also requires production of EPS. Vps refers to the EPS produced by *V. cholerae*. Figure and legend reproduced from Davey and O'Toole (2000).

Watnick and Kolter (1999) and Pratt and Kolter (1998) report similar roles for flagella, pili and EPS in *Vibrio cholerae* and *E. coli* biofilms respectively, while Deziel *et al.* (2001) show that the

overuse of these structures in biofilm-forming strains reduces motility and chemotactic responses. A link between motility and biofilm formation has also been shown in strains of *P. fluorescens* (Korber *et al.*, 1994; Casaz *et al.*, 2001). This is interesting considering that WspR, shown to play a role in cellulose synthesis in this study, is homologous to PleD which regulates flagellar degradation in *C. crescentus*.

The biofilm formed in the LSWS genotype of *P. fluorescens* is composed of cellulose (Spiers *et al.*, 2002) and a calcofluor stain to demonstrate this is shown in Figure 1.6. This is not necessarily the main component of biofilms in other species. In *P. aeruginosa* (which has no *wss* operon homologue) an uncharacterised exopolysaccharide or lipopolysaccharide is responsible, whereas the proteinaceous substance curli is involved in the case of *E. coli* (Olsen *et al.*, 1989).

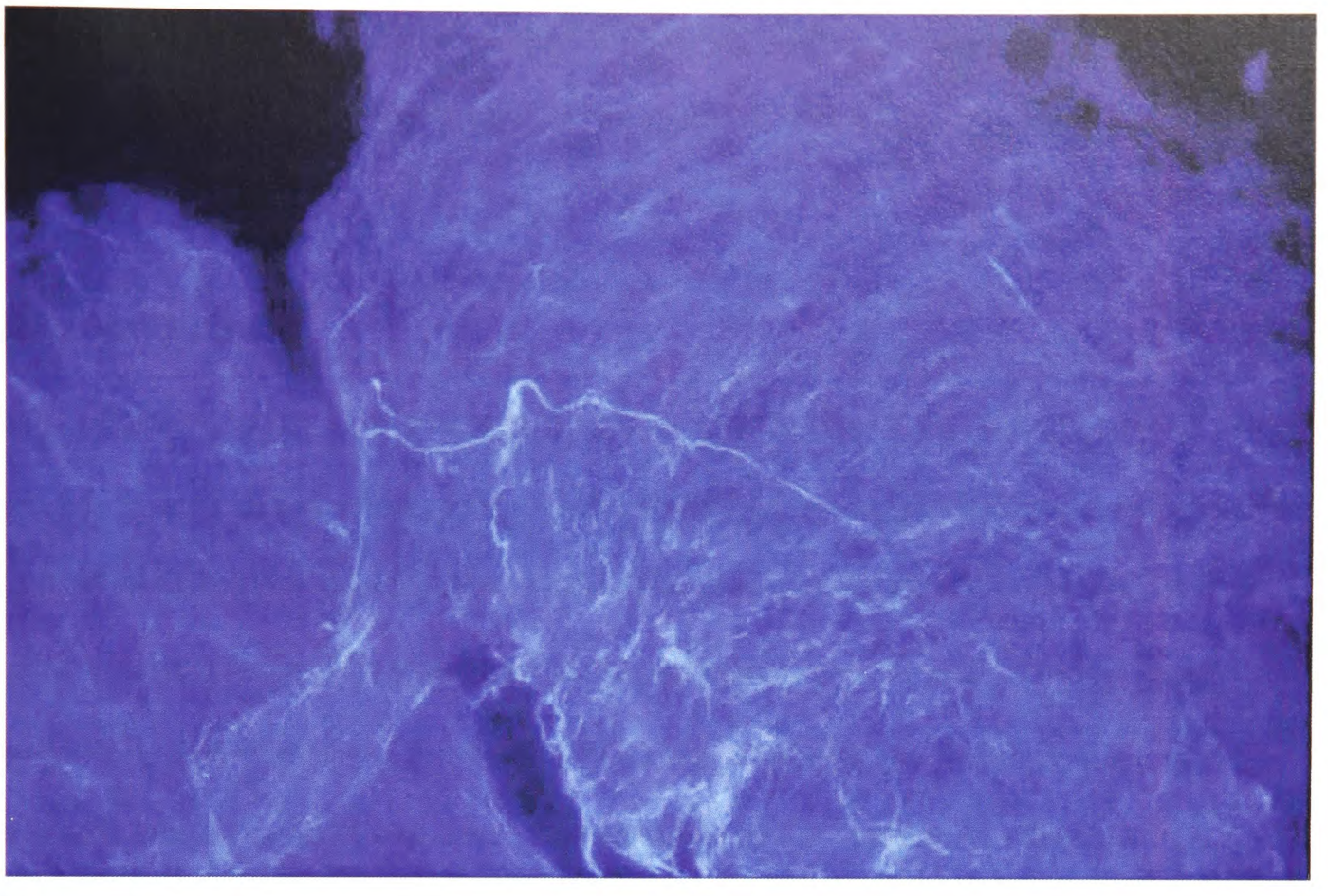


Figure 1.6. WS biofilm stained with 10μM Calcofluor (UV microscopy). The network composed of a cellulose-like polymer (CLP) is visible as a series of nets and irregular fibres. Figure obtained from A. Spiers (personal communication).

Genetic approaches to the study of biofilm formation (reviewed in O'Toole *et al.*, 1999, and Pratt and Kolter, 1999) have uncovered many regulatory pathways that play a role. In *E. coli* the *cpx* pathway (Dorel *et al.*, 1999) and the *ompR* gene (Vidal *et al.*, 1998) have both been implicated and Rashid *et al.* (2000) have shown the involvement of the polyphosphate kinase genes in *P. aeruginosa*. O'Toole and Kolter (1998b) conclude that multiple signalling pathways are involved in *P. fluorescens*. In most of these cases the wild-type strain is able to form a biofilm in a phenotypically plastic way, and a mutant screen has been carried out to look for deficient genotypes. In the adaptive radiation of *P. fluorescens* SBW25 an obligate biofilm-former evolves from an ancestor that does not appear to form a biofilm, at least in the environment studied. In this respect it is similar to the *E. coli* case described by Vidal *et al.* (1998) where an *ompR* mutation produces a biofilm-former, though Deziel *et al.* (2001) attribute such a case in *P.*

aeruginosa to phase variation. These examples show the importance of regulatory genes in the evolution of biofilm formation, irrespective of the chemical structure of the biofilm itself.

1.10 The *wsp* operon

The WS4 mutation, which blocks formation of the LSWS phenotype, disrupts *wspR*. This is, however, the seventh and terminal gene of the *wsp* operon, in which transposon mutants WS7, WS11 and WS29 were also found (Kahn, 1998). This does not necessarily implicate the other six genes of the operon in the WS phenotype as the mutations may be polar, but the genes in the operon are likely to act together. Bantinaki (2001) has performed an extensive study of the *wsp* operon which will be summarised here. The structure of the operon is shown in Figure 1.7, including the closest homologues of the seven genes. The operon is arranged such that the genes *wspA* to *wspE* overlap and thus may be translationally coupled, whereas *wspF* and *wspR* are separated from the other genes by about 50bp.

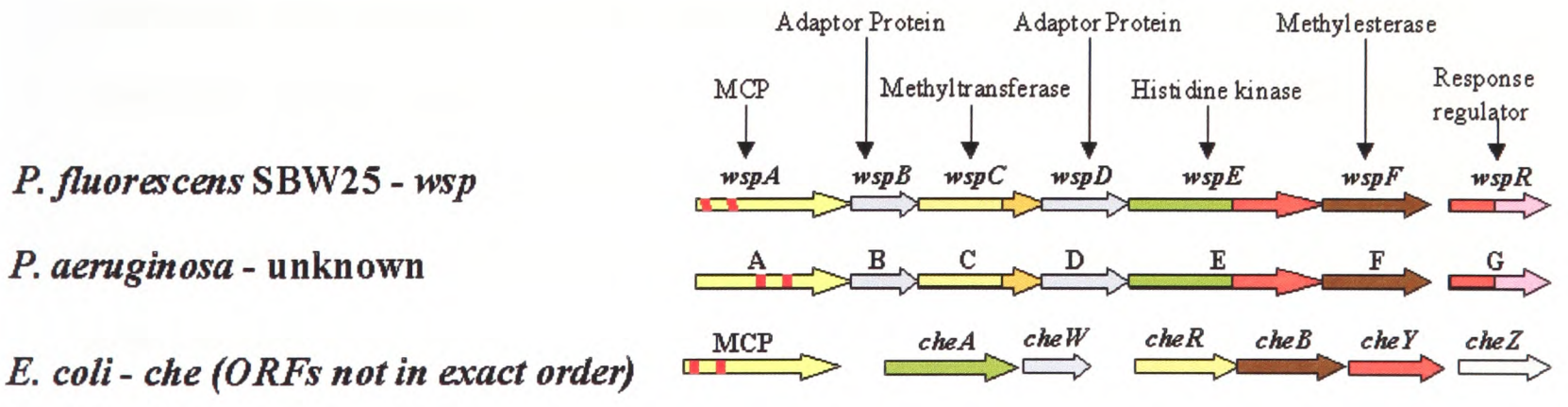


Figure 1.7. Structure of the *wsp* operon. The *wsp* operon of *P. fluorescens* is shown with functional annotation. Also shown are the homologous operons from *P. aeruginosa* and *E. coli*. Homologous domains are represented in the same colour, and the transmembrane helices of the MCPs are marked in maroon. Figure obtained from E. Bantinaki (personal communication).

The first gene, *wspA*, encodes a methyl-accepting chemotaxis protein, bearing closest relationship to MCP-2 of *E. coli*. This is a membrane-bound protein, whose periplasmic domain binds

signalling ligands, while its cytoplasmic domain interacts with a CheA-like protein. Functional sequence analysis shows that the N terminus of WspA is cytoplasmic and the first domain is a membrane-spanning one. This is followed by the periplasmic ligand-binding domain; these are highly variable in nature, which is probably indicative of the wide range of ligands that function as signalling molecules. The next domain is another membrane-spanning one; the two membrane-spanning domains must in some way transmit the structural change to the cytoplasmic domain that occur on ligand binding to the periplasmic domain, so that the former becomes methylated and hence interacts with the histidine kinase, CheA. The C-terminal domain contains a 61-residue conserved motif thought to be the site of interaction with CheA and the copies of the scaffold protein CheW. Also in the cytoplasmic domain are the glutamate and glutamine residues that are methylated in response to ligand binding.

The two genes, *wspB* and *wspD*, encode homologues to the single domain CheW proteins of *E. coli* and *Rhizobium* species. These are adaptor, or scaffolding, proteins and are not known to possess any catalytic function. They have, however, along with CheA, been implicated in the localisation and clustering of MCPs (Maddock and Shapiro, 1993). It is not clear why a chemotactic operon should require two *cheW* genes; possibilities include different growth conditions requiring different scaffolds or the interaction with other MCPs resulting in cross-talk between pathways. This latter case is of particular evolutionary interest as it might be the molecular basis of exaptations.

The third gene in the operon, *wsp*C, is a homologue of *cheR* from *E. coli* and *frzF* from *Myxococcus xanthus*. The products they encode are responsible for the methylation of glutamate residues in the MCP. CheR possesses a catalytic N terminus and multiple C terminal domains involved in the protein-protein interaction. FrzF, however, in common with WspC, does not possess these C terminal domains so must interact with the MCP in an uncharacterised manner. In

its place it has tetratrico peptide repeat (TPR) motif which is a 34-residue sequence that is involved in protein-protein interactions in a wide variety of systems and organisms (Lamb *et al.*, 1995). Up until now it has not been found in chemotactic systems.

The *wsp*E gene product is a hybrid protein that shows homology to parts of both the histidine kinase, CheA and response regulator, CheY. Its closest homologue is FrzE from *M. xanthus*, to which it is homologous across its length, but the *che* operon of *E. coli* has been more highly studied and provides more functional insights (e.g. Aizawa *et al.*, 2000). CheA is the histidine kinase that autophosphorylates in response to the methylation of the MCP, and then phosphorylates the response regulator (CheY in the *che* system; WspR in the *wsp* system). CheA has four domains with the following functions: autophosphorylation; kinase activity; receptor binding; CheY binding. Of these, WspE and FrzE possess the first three, but lack the CheY-binding domain. This can perhaps be explained by the fact that the other section of the hybrid is itself a CheY homologue, containing all the conserved response regulator elements. It is thought that some kind of phosphotransfer mechanism occurs between this domain and the kinase. The function of WspE is not entirely clear; neither is the possession of two CheY domains in the operon: as shall be discussed below WspR is a hybrid protein also possessing a CheY domain. There is a correlation between the lack of a *cheZ* homologue and the presence of multiple *cheY* homologues so the proteins encoded by the latter may be acting as phosphate sinks.

WspF is a methylesterase with homology to CheB. The latter is responsible, along with CheR, for receptor methylation: CheR catalyses methylation, while CheB catalyses the opposite reaction. The protein consists of two domains, an input one that is phosphorylated and an output one that possesses catalytic activity. In this it is similar to WspR which is also made up of two domains. It will be discussed at greater length below but briefly it possesses an input, CheY-like, domain, and an output domain of unknown function containing the GGDEF motif.

The mechanism of action of the entire *che* operon is illustrated in Figure 1.8. This is the most well studied operon [for a review see Falke *et al.* (1997)] but is thought to be similar to others such as the *frz* operon of *M. xanthus* (Kearns and Shimkets, 2001). Bantinaki (2001) demonstrated the similarity of the *che* and *wsp* operons.

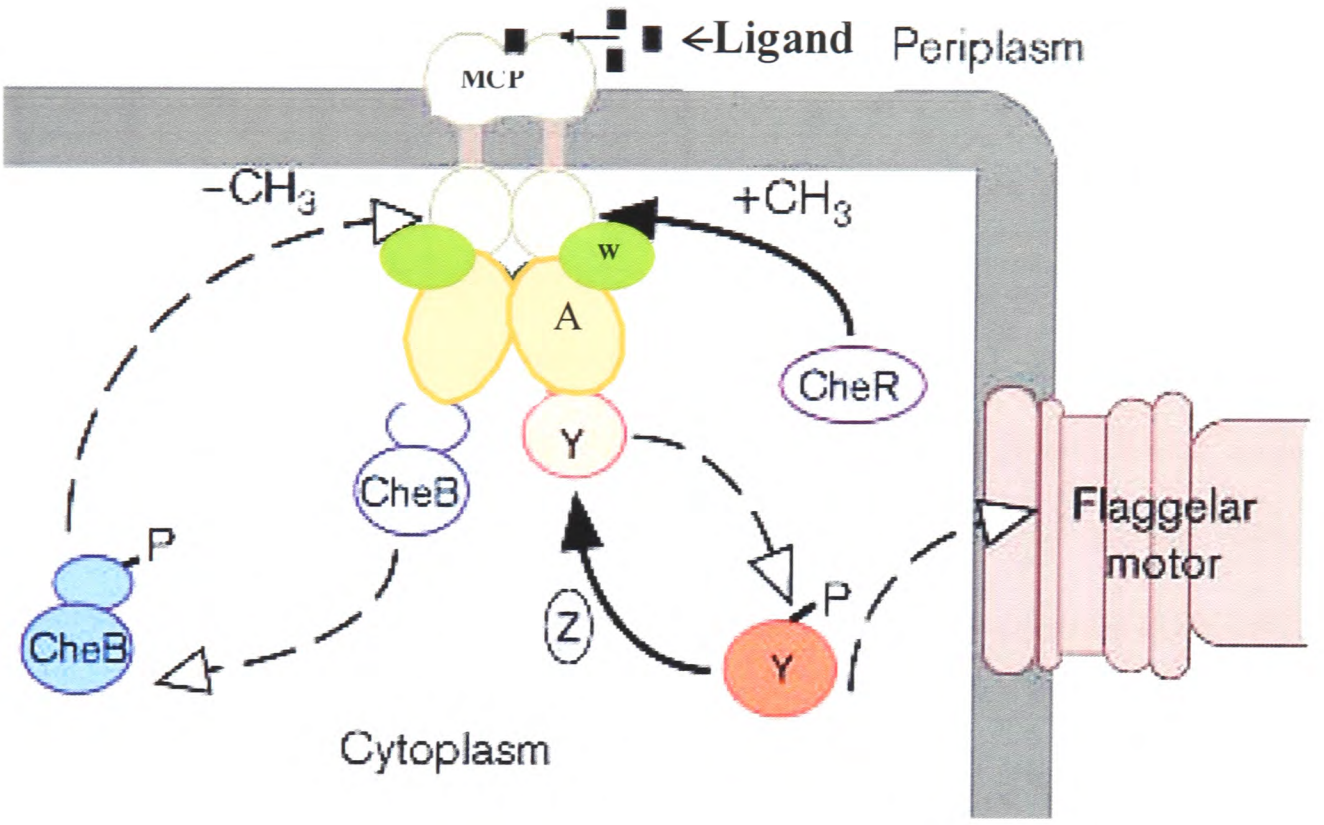


Figure 1.8. The chemotaxis pathway. Unbound and/or highly methylated chemoreceptor (MCP) activates the kinase CheA to transfer a phosphate to the response regulators CheY and CheB. High levels of phosphorylated CheY increase the frequency of switching to clockwise flagellar rotation, thus increasing tumbling. Active phosphorylated CheB slowly demethylates the receptor to reset the signalling state. When a receptor binds ligand and/or is unmethylated, CheA is inactive, resulting in reduced levels of phosphorylated CheY, and therefore, more counterclockwise flagellar rotation and more running (smooth swimming). With CheB also inactive, the methyltransferase activity of CheR decreases receptor sensitivity. Constitutively activated forms of the proteins are shown by continuous arrows. CheW is the adaptor protein. Figure and legend reproduced from Bantinaki (2001).

The *wsp* operon has two known homologues at the entire operon level: an operon from the entire genome sequence of *P. aeruginosa* PA01 (Stover *et al.*, 2000) and one from the megaplasmid pMLb of *Mesorhizobium loti* (Kaneko *et al.*, 2000). The function of these operons is unknown and *P. aeruginosa* lacks a cellulose synthase operon so the operon is either redundant or playing a different role. This raises the possibility that cellulose production is not the only role of the *wsp* operon in *P. fluorescens*.

1.11. Two-component signal transduction

The discovery that *wspR*, identified in a mutant screen for the WS cellulose-synthesising phenotype, encodes a response regulator at the end of a chemosensory-like operon necessitates a discussion of bacterial signal transduction. Signal transduction pathways couple external stimuli, such as nutrient availability or stress, with appropriate responses such as motility or gene expression. This is done through a defined series of protein-protein interactions, modifications and small second messenger molecules. Prokaryotic and eukaryotic signal transduction share many features and indeed homologies, but there is a quantitative difference with one type of system predominating in the former, and a different one in the latter (Stock *et al.* (2000) review the taxonomic distribution). In eukaryotes the serine/threonine and tyrosine kinase systems, that create phosphoesters, predominate, whereas in bacteria the histidine kinase system, of which the *che* and *wsp* pathways are examples, that creates greater negative free energy phosphoramidates and acyl phosphates, is more prevalent. This bacterial system is referred to as two-component signal transduction.

This two-component system has been widely studied and reviewed, though the nomenclature varies: Egger *et al.* (1997) refer to it as the histidyl-aspartyl phosphorelay, while Hoch (2000) uses it in a narrower sense to distinguish it from more complicated phosphorelay systems; both varieties will be discussed here. The essential two components of the system are the histidine kinase (HK) sensor protein (such as CheA and WspE), and the response regulator (RR; such as CheY and WspR). The former typically consists of two domains, a sensor domain that varies widely and a conserved C-terminal catalytic domain that autophosphorylates in response to a signal at a histidine residue, on the N3 of the imidazole ring. This phosphate is transferred to a conserved aspartate residue (typically aspartate-57) in the N terminus of the response regulator, a

reaction catalysed by that regulator itself. This causes activation of the C terminal effector domain which, in most cases studied thus far, is a DNA-binding element. Finally the phosphate is removed, either by a phosphatase or by autophosphatase activity of the response regulator itself. The chemistry of the pathway can be summarised thus:

1. Autophosphorylation: $HK\text{-}His + ATP \leftrightarrow HK\text{-}His\sim P + ADP$
2. Phosphotransfer: $HK\text{-}His\sim P + RR\text{-}Asp \leftrightarrow HK\text{-}His + RR\text{-}Asp\sim P$
3. Dephosphorylation: $RR\text{-}Asp\sim P + H_2O \rightarrow RR\text{-}Asp + P_i$

These reactions are dependent on the presence of divalent cations, magnesium fulfilling that role *in vivo*.

While this is the basic outline of the system there is considerable variation in nature, in both the structures of the components and their interactions. The classical type of HK is membrane bound, with the extracellular N terminus receiving the signal to promote the formation of homodimers. Autophosphorylation occurs between the C termini of the two components of the dimer; an example of this class is the EnvZ HK of *E. coli* (Tanaka *et al.*, 1998; Tomomori *et al.*, 1999). However, not all HKs are membrane bound: CheA, as mentioned above (also Bilwes *et al.*, 1999), is a cytoplasmic protein that interacts with a membrane-bound receptor and other HKs interact with small cytosolic molecules; the variety of such receptors is reviewed by Alexandre and Zhulin (2001). Some HKs are complex phosphorelay systems such as ArcB of *E. coli* (Ishige *et al.*, 1994). This contains two kinase domains and an aspartate receptor domain and the phosphate signal passes through all three before reaching the RR (see Figure 1.9b). In other systems such as sporulation control in *Bacillus subtilis* (Hoch, 1993; Perego, 1998) the corresponding kinases and aspartate-domains of the phosphorelay are separate proteins (see Figure 1.9d). While the response regulator is normally a DNA-binding protein, other forms are

found. CheY, as described above (see also Stock *et al.*, 1989), is a single domain protein that interacts with components of the flagella, while CheB (Simms *et al.*, 1985) has an enzymatic effector domain. The operation of these examples of two-component systems is illustrated in Figure 1.9.

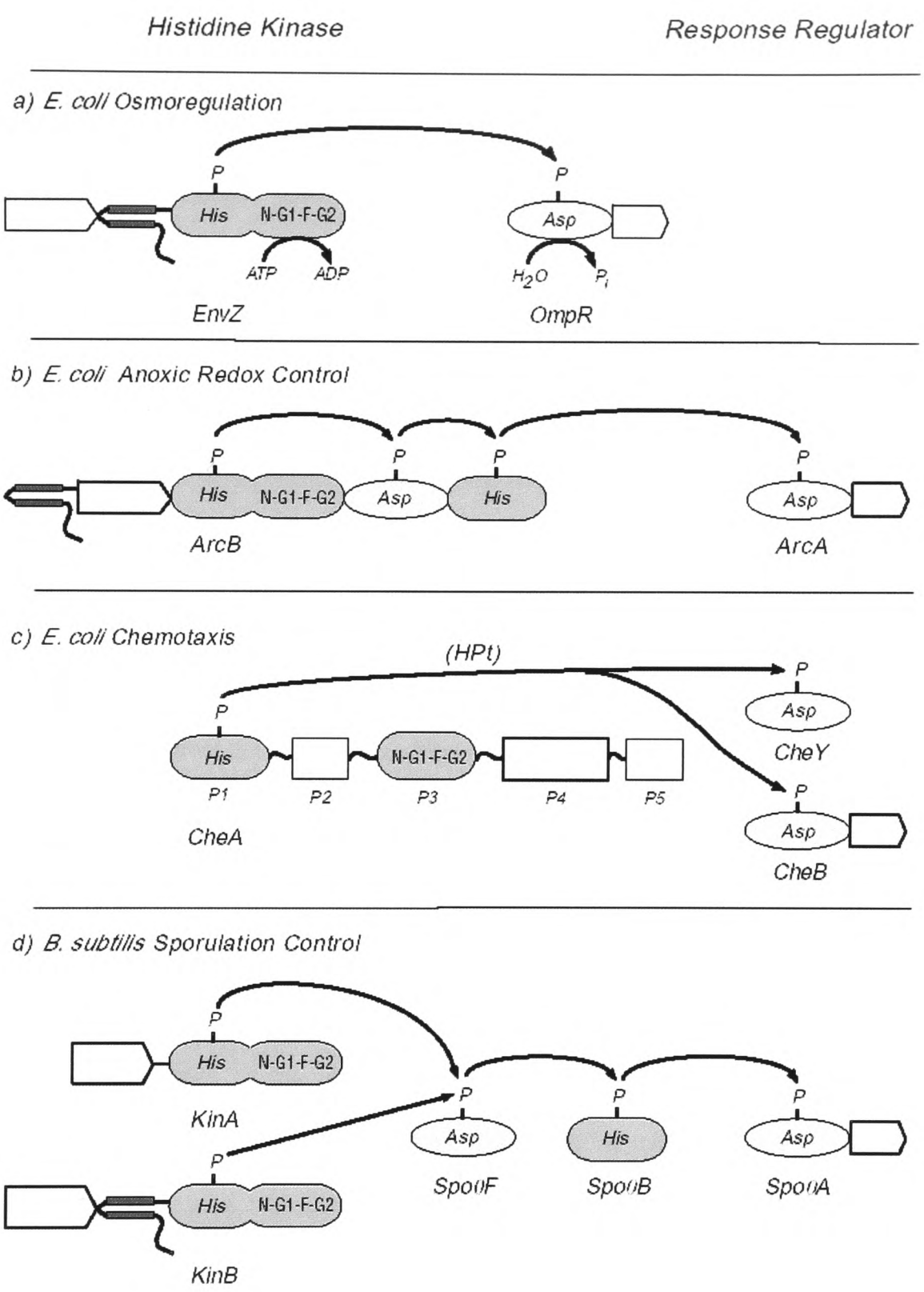


Figure 1.9. Schematic diagram depicting the modular organisation of representative two-component systems. Asp-containing domains are coloured dark grey, His-containing domains are coloured light grey, and variable auxiliary domains are coloured white. (a) The prototypical two-component pathway exemplified by the *E. coli* osmoregulatory system uses a single phosphoryl transfer event between the orthodox histidine protein kinase (HK) EnvZ and its cognate response regulator protein (RR) OmpR. (b) The *E. coli* Arc system illustrates a phosphorelay involving the hybrid HK ArcB. Depending on aerobic conditions, ArcA is capable of receiving a phosphoryl group from either the catalytic core or the His-containing phosphotransfer (HPt) domain of ArcB. (c) The *E. coli* chemotaxis pathway involves an atypical soluble HK CheA that phosphorylates either of two RRs, the single domain RR CheY and the methylesterase CheB. (d) The *B. subtilis* sporulation control system is a multicomponent His-Asp-His-Asp phosphorelay system in which all of the signalling domains are independent proteins. Spo0F receives a phosphoryl moiety from either KinA or KinB and subsequently transfers it to the HPt Spo0B, which then phosphorylates the terminal RR Spo0A. Figure and legend reproduced from Stock *et al.* (2000).

Several studies have the shown the effects of mutations in bacterial signal transduction systems; these show that selection could potentially act here to create new adaptive phenotypes. Tran *et al.* (2000) show that the regulation of *ompC* and *ompF* by OmpR in *E. coli* can be altered by single amino acid substitutions. These affect the relative rates of transcription of the two genes in response to osmotic changes. Han *et al.* (1997) show that a spontaneous duplication occurs in the PheN response regulator of *Pseudomonas tolaasii* to alter pathogenicity and colony morphology. Brown *et al.* (2001) demonstrated that upregulation of the novel response regulator MlrA in both *E. coli* and *S. typhimurium* can increase curli production and aggregative behaviour, while Barak and Eisenbach (1999) show that *cheY*, when overexpressed in the absence of the other *che* genes, can still produce a chemotactic response, implying cross-talk with other pathways. The evolved WS morph, through the action of WspR, may have similarities with some of these systems.

1.12 WspR and its homologues

The seventh and final gene in the *wsp* operon, *wspR*, encodes a response regulator of the type described above. It has been characterised by Kahn (1998) and Bantinaki (2001) and will be further examined in this study, including an analysis of its domain structure in Chapter 3. One of its closest, and most characterised, known relatives is the *pleD* gene from *C. crescentus*. This gene is involved in the cell-cycle dependent control of flagellar degradation (Sommer and Newton, 1989; Hecht and Newton, 1995; Aldridge and Jenal, 1999). The cell cycle of *C. crescentus* is illustrated in Figure 1.10 and its control is reviewed by Jenal (2000).

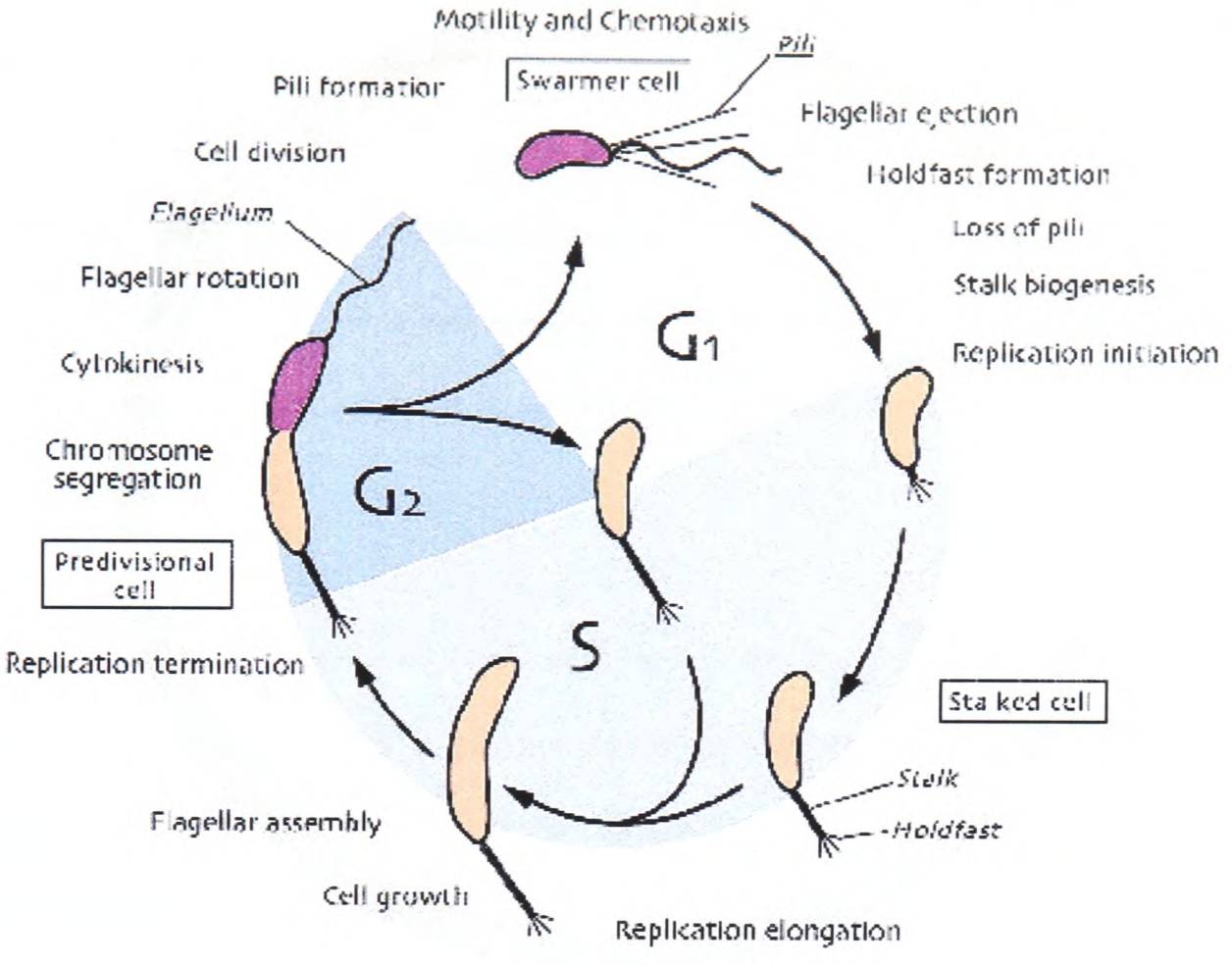


Figure 1.10. Schematic diagram of the *C. crescentus* cell cycle. The three different cell types of *C. crescentus*, the stalked cell, the swarmer cell, and the predivisional cell are depicted with polar appendices underlined. The clockwise orientation of the cell cycle is indicated by arrows. The developmental and cell cycle events described in the text are highlighted at the time when they take place during the life cycle. The *Caulobacter* cell cycle is divided into three distinct phases: The G1-phase includes the motile stage and the swarmer-to-stalked cell transition; the S-phase defines the period of active chromosome replication ; the G2-phase outlines cytokinesis and cell division stages. The length of the S-phase is identical for both the newly differentiated (inexperienced) stalked cell and the (experienced) stalked cell emerging from cell division. Figure and legend reproduced from Jenal (2000).

The essential feature of the *C. crescentus* cell cycle is that it involves two different cell types: the swarmer cell and the stalked cell, of which only the latter is capable of replication. Cell division is asymmetric, producing one of each type of daughter cell. The sessile stalked cell divides again immediately, while the motile swarmer cell may move location by chemotaxis before it begins the transition to becoming a stalked cell. This transition requires the ejection of the flagellum which is where PleD is involved. Sommer and Newton (1989) identified several mutants that were cell-cycle deficient. Amongst these, *pleC* was non-motile and *pleD* was supermotile. Burton *et al.* (1997) showed that PleC is a histidine kinase that regulates the onset of motility. Hecht and

Newton (1995) showed that the *pleD* mutation is a suppressor of *pleC*, and is in a gene encoding a novel response regulator that is required for flagellar ejection. Aldridge and Jenal (1999) showed that this was due to an interaction between PleD and the flagellar motor component protein FliF. WspR and PleD are homologous across two domains, though the similarity is greater in the C terminus and PleD possesses an additional N terminal domain. This will be discussed further in Chapters 3 and 4.

Another gene that has similarity to *wspR* and *pleD* is the *hmsT* gene of the plague bacteria, *Yersinia pestis* (Jones *et al.*, 1999). The Hms+ phenotype of this organism is defined as the ability to colonise and block the proventiculus of the oriental flea rat, *Xenopsylla cheopis*, which results in repeated attempts to feed and hence successful infection of the mammalian host. The structural operon responsible for this phenotype is *hmsHFRS* (Pendrak and Perry, 1991), but Jones *et al.*, (1999) show that another, regulatory, locus is also required. This is the *hmsT* locus that encodes the PleD relative. Nothing is known, however, about its mechanism of action.

As has already been mentioned above, CelR2 from *R. leguminosarum* (Ausmees *et al.*, 1999) and AdrA from *S. typhimurium* (Romling *et al.*, 2000; Zogaj *et al.*, 2001) are also related to PleD and WspR (and CelR2 and WspR are homologous across their entire lengths) and in these cases they are also involved in cellulose biosynthesis. The most conserved feature amongst all these proteins is the C terminal GGEEF or GGDEF motif, which has been found in the genomes of many bacteria that have been the subjects of sequencing projects, and has received particular attention as one of the hypothetical proteins of unassigned function (Galperin, 2001). The motif is found as both GGDEF and GGEEF (its sequence in WspR), but all domains that contain it are named GGDEF domains. The genomic and taxonomic distribution of this domain amongst the bacteria is discussed by Galperin *et al.* (2001) and it is seen to be widespread in many different putative

signalling proteins, appearing, for example, 32 times in *M. loti*, 33 in *P. aeruginosa*, 19 in *E. coli* and 11 in *C. crescentus*, but so far not observed in archaea or eukaryotes.

1.13 Cyclic di-GMP

There are few clues as to the function of GGDEF proteins, the study of PleD (Aldridge and Jenal, 1999) being one. The other line of approach, that is potentially more relevant to WspR, is the potential interaction with cyclic di-GMP. In *A. xylinum* two proteins are important in the turnover of this particular nucleotide: diguanylate cyclase (DGC) that catalyses its formation from the linear dinucleotide, and phosphodiesterase A (PDEA) that degrades it by cleavage to create a different linear dinucleotide (Tal *et al.*, 1998). Both these enzymes (which each have three isoforms) possess GGDEF domains in their C termini, suggesting a critical role for this motif in the interaction with cyclic di-GMP, though both proteins also possess EAL domains. Further evidence for the role of the GGDEF domain comes from Pei and Grishin (2001) who use a bioinformatic approach to show homology at the three-dimensional protein level between the GGDEF motif and the catalytic domain of adenylyl cyclase. Notably, Ausmees *et al.* (2001) show functional complementation between DGC from *A. xylinum*, CelR2 from *Rhizobium*, and an uncharacterised GGDEF protein from *E. coli*.

The reason that cyclic di-GMP is of particular interest in this case is because of the role it plays as a regulator of cellulose synthesis (Ross *et al.*, 1987; 1990). It has been shown that this effect is mediated by the c-di-GMP-binding protein, that interacts with the cellulose synthase complex (Weinhouse *et al.*, 1997). Similar systems have been found in *A. tumefaciens* (Amikam and Benziman, 1989) and plants (Amor *et al.*, 1991). Therefore it is thought that WspR may affect cellulose synthesis through c-di-GMP. Interestingly it was the localisation of this c-di-GMP-binding protein at the bacterial poles that lead to the finding that cellulose synthesis is thus

located (Kimura *et al.*, 2001). Protein localisation in bacteria (reviewed by Lybarger and Maddock, 2001) involves both structural genes such as cellulose synthase and regulatory ones (Harrison *et al.*, 1999; Boyd, 2000). The localisation of cellulose synthesis has been demonstrated by fluorescence microscopy in the WS morph of *P. fluorescens* (Bohannon, White and Rainey, unpublished data) and it would be interesting to know whether this also applies to any of the regulatory proteins, such as WspR.

These studies of *wspR* homologues point very strongly to the involvement of WspR in *wss* operon function through cyclic di-GMP, rather than through the more usual response regulator function of transcriptional activation. This will be investigated more thoroughly in this study.

1.14 The candidate gene approach

Having identified the gene of interest for this study (*wspR*), it should be put back into its evolutionary context with an analysis of the candidate gene approach. Palopoli and Patel (1996) assess the usefulness of this approach by returning to the issues of micro- and macromutationism. It has been noted by many developmental geneticists that mutations obtained in the laboratory at candidate loci, such as *wspR*, often produce phenotypes with striking resemblance to adaptations in related species and taxa (or in this case genotypes). When coupled with studies such as that by Schmidt *et al.* (2000), which show adaptive genetic variation at such loci between species, it is appealing to speculate that these types of mutations typify the genetics of adaptation. This line of thinking has been attractive to authors such as Gould (1977), but less so to micromutationists such as Mayr and Provine (1980) who argue that any mutation that is large enough to be visible to the experimenter is too big to be advantageous. More recent authors have suggested that the truth is somewhere in between and more importantly is open to experimental test. Such tests involve the artificial manipulation of phenotype by mutations in the candidate loci, and subsequent

assessment of the resultant fitness. This can be achieved by the use of transgenes (Tatar, 1999; Cavener, 1992), with either artificial mutations being introduced, or genes from a related, but differently adapted, species. When genes are moved between species to test for adaptation, they may assist in the distinguishing of *cis* from *trans* factors involved in regulation. If the activity of the transgene is responsible for the adaptation, its ability to complement in the recipient species will only occur if the necessary regulatory elements are in *cis*. Palopoli and Patel (1996) suggest that an important distinction exists between mutations affecting candidate loci in *cis* and those whose effect is in *trans*. The former may be of moderately large effect while the latter are likely to be very small alterations to a complicated regulatory network due to pleiotropy. Haag and True (2001) suggest that phylogenetic distance may be important, with large effects at candidate loci being relevant between closely related species but becoming less important with increased divergence in the developmental genetic regulatory systems involved. The evidence therefore suggests that the candidate gene approach, that looks for adaptive mutations at particular loci, is useful, albeit with several caveats.

If this is the case, then the manipulation of both fitness and phenotype with transgenes is a powerful approach even within a species, and the engineering of *hsp70* in *D. melanogaster* is a significant example (Krebs and Feder, 1998; Feder and Krebs, 1998; reviewed by Feder, 1999). In this case the fitness of the flies at different stages in their life cycles and at different temperatures was manipulated by overexpression and mutation of Hsp70. This allowed the analysis of trade-offs in fitness components and the role of this particular gene. The experimental approach (reviewed by Schmitt, 1999) is not limited to genetics; the use of hormonal and other phenotypic manipulation can also prove insightful (Hegner and Wingfield, 1987; Ketterson and Nolan, 1999; Sinervo, 1999) in assessing selection.

The candidate gene approach is however one step closer to 'finding the genes that matter', than the non-genetic methods. Feder (1999) rightfully points out, though, that this approach should be just one component of a multidisciplinary effort. The other elements are the studies of developmental genetics, population genetics, genomics, and random mutagenesis. That author argues that these approaches are rarely all possible in the same organism, some being confined to model organisms lacking a natural environment, and others being quite the opposite. The system used in this study has advantages over many in this respect. As a bacterium the genetics are more tractable than most eukaryotes and a sequenced genome is on the horizon (P.B. Rainey, personal communication). Also the evolutionarily relevant environment in this study is the laboratory microcosm. Hence *wspR*, identified by random mutagenesis, is an ideal candidate gene. Such an approach will be used extensively in Chapters 5 and 6.

1.15 Study Aims.

1. Phenotypic characterisation of *wspR* point mutants.
2. Elucidation of the level of action of WspR.
3. Directed mutagenesis and use of homologues to probe WspR structure-function relationship.
4. Manipulation of fitness using *wspR* alleles.
5. Assessment of the correlation between phenotype and fitness.
6. Study of differences between independently isolated wrinkly spreaders.
7. Understanding of the role of *wspR* in different wrinkly spreaders and its involvement in the differences between them.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

Strain	Description	Reference
Pseudomonas fluorescens		
SBW25 SM	Environmental isolate; wild-type ancestral smooth genotype	Rainey and Bailey (1996)
SBW25 PR1200 (or LSWS)	Biofilm-former, evolved from SBW25 SM (Large Spreading Wrinkly Spreader)	Spiers *et al.* (2002)
SBW25 WS4	LSWS *wspR*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 WS7	LSWS *wspE*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 WS11	LSWS *wspE*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 WS13	LSWS *wspE*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 WS29	LSWS *wspC*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 WS40	LSWS *wspC*::miniTn*5*, Km^R	Spiers *et al.* (2002)
SBW25 NM	non-motile isolate evolved from SBW25 SM	Rainey Lab stock
SBW25 AS14	SBW25 SM with *wssE::lacZ* fusion, Tc^R	Rainey Lab stock
SBW25 AS15	SBW25 LSWS with *wssE::lacZ* fusion ,Tc^R	Rainey Lab stock
SBW25 WS a-z	26 independent WS isolates from SBW25 SM	Kahn (1998)
SBW25 AS17	WS e with *wssE::lacZ* fusion Tc^R	Rainey Lab stock
SBW25 AS18	WS i with *wssE::lacZ* fusion Tc^R	Rainey Lab stock
SBW25 AS19	WS n with *wssE::lacZ* fusion Tc^R	Rainey Lab stock
SBW25 AS20	WS s with *wssE::lacZ* fusion Tc^R	Rainey Lab stock

SBW25 AS21	WS v with *wssE::lacZ* fusion Tc^R	Rainey Lab stock
SBW25 AS22	WS x with *wssE::lacZ* fusion Tc^R	Rainey Lab stock
SBW25 LSWS Δ*wsp*	LSWS with deletion of *wspA-F*	Bantinaki (2001)
SBW25 SM Δ*wsp*	SM with deletion of *wspA-F*	Bantinaki (2001)
SBW25 SM*panB*	independent pantothenate mutant of SM	Rainey and Travisano (1998)
SBW25 WS*panB*	independent WS pantothenate mutant of SM	this study
SBW25 JB01	SBW25 with NPTII promoter for *wss* operon	Rainey Lab Stock
Caulobacter crescentus		
UJ284	CB15, syn-1000, Δ*pleD*	Aldridge and Jenal (1999)
Escherichia coli K12		
S17-1 *λpir*	C600, *recA, thi, pro, thr, leu, sulI, hsdm*$^+$, lysogenised with phage *λpir*	Simon *et al.* (1983)
DH5α	*endA*1, *hsdR*17(r_K-m_K+), *supE*44, *recA*1, *gyrA* (Nal^r), *relA*1, Δ(*lacIZYA-argF*)U169, *deoR*, ϕ80*dlacΔ(lacZ)M15*	Woodcock *et al.* (1989)

2.1.2 Plasmids

Plasmid	**Description**	**Reference**
pVSP61	Km^R, *lacZ*, 13.5kb *Pseudomonas* expression vector	Loper and Lindow (1994)
pVSP61-*wspR9*	pVSP61 with *wspR*9 as *Bam*H1-*Eco*R1 fragment	Kahn (1998)
pVSP61-*wspR19*	pVSP61 with *wspR*19 as *Bam*H1-*Eco*R1 fragment	Kahn (1998)
pVSP61-*wspR13*	pVSP61 with *wspR*13 as *Bam*H1-*Eco*R1 fragment	Kahn (1998)
pVSP61-*wspR14*	pVSP61 with *wspR*14 as *Bam*H1-*Eco*R1 fragment	Kahn (1998)
pVSP61-*wspR5*	pVSP61 with *wspR5* as *Bam*H1-*Eco*R1 fragment	Kahn (1998)

pVSP61-*wspR12*	pVSP61 with *wspR* as *Bam*H1-*Eco*R1 fragment	Kahn (1998)
pVSP61-*wspR1*	pVSP61 with *wspR*1 as *Bam*H1-*Eco*R1 fragment	this study
pHP45-ΩTc^R	pBR-type plasmid containing omega Tc^R fragment	Fellay *et al.* (1989)
pVSP61-*wspR9*-ΩTc^R	pVSP61-*wspR9* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-*wspR19*-ΩTc^R	pVSP61-*wspR19* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-*wspR13*-ΩTc^R	pVSP61-*wspR13* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-*wspR14*-ΩTc^R	pVSP61-*wspR14* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-*wspR5*-ΩTc^R	pVSP61-*wspR5* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-*wspR12*-ΩTc^R	pVSP61-*wspR12* with ΩTc^R cloned into *EcoR1* site	this study
pVSP61-ΩTc^R	pVSP61 with ΩTc^R cloned into *EcoR1* site	this study
pRK2013	IncP4, *tra*, *mob*, Km^R, helper for triparental mating	Ditta *et al.* (1980)
pVSP61-*wspR*NtermA	pVSP61 with *wspR* N term (version A) as *BamH1-EcoR1* fragment	this study
pVSP61-*wspR*NtermB	pVSP61 with *wspR* N term (version B) as *BamH1-EcoR1* fragment	this study
pVSP61-*wspR*Cterm	pVSP61 with *wspR* C term as *BamH1-EcoR1* fragment	this study
pVSP61-*wspR12d*	pVSP61 with *wspR* allele 12d as *BamH1-EcoR1* fragment	this study
pVSP61-*wspR19d*	pVSP61 with *wspR* allele 19d as *BamH1-EcoR1* fragment	this study
pVSP61-*wspR4*	pVSP61 with *wspR* allele 4 as *BamH1-EcoR1* fragment	this study
pMR20	RK2-based broad-range expression vector, Tc^R	Aldridge and Jenal (1999)
pPA13	pBluescriptSK containing *pleD*-GG368DE, Ap^R	Aldridge and Jenal (1999)
pPA41	pMR20 expressing *pleD*; Tc^R	Aldridge and Jenal (1999)
pPA99	pMR20 expressing *pleD-wspR9*; Tc^R	P. Aldridge*

pPA100	pMR20 expressing *pleD-wspR12*; Tc^R	P. Aldridge*
pPA110	pMR20 expressing *pleD*-GG368DE-*wspR12*; Tc^R	P. Aldridge*
pPA11428	pMR20 expressing *pleD*1*; Tc^R	P. Aldridge*
pPA11433	pMR20 expressing *pleD*2*; Tc^R	P. Aldridge*
pPA11435	pMR20 expressing *pleD*3*; Tc^R	P. Aldridge*
pPA11441	pMR20 expressing *pleD*4*; Tc^R	P. Aldridge*
pPA11443	pMR20 expressing *pleD*5*; Tc^R	P. Aldridge*
pPA11447	pMR20 expressing *pleD*6*; Tc^R	P. Aldridge*
pVSP61-*wspRpleD*1	pVSP61 with *wspR-pleD* hybrid 1 as *Eco*R1-*Bam*H1 fragment	this study
pVSP61-*wspRpleD*2	pVSP61 with *wspR-pleD* hybrid 2 as *Eco*R1-*Bam*H1 fragment	this study

*These plasmids were the kind gift of Phillip Aldridge, University of Basel.

2.1.3 Media and culture conditions

Bacteria were cultured in Luria-Bertani medium (LB. 10g NaCl; 10g tryptone; 10g yeast extract (BDH) L^{-1}), salt-free LB (LBns. 10g tryptone; 10g yeast extract (BDH) L^{-1}), *Pseudomonas* agar F (PAF. 10g glycerol; 10g tryptone; 10g Protease Peptone No.3 (BDH); 1.5g $Mg_2SO_4.7H_2O$; 1.5g $K_2HPO_4.3H_2O$ L^{-1}), King's medium B (KB. 10g glycerol; 20g Protease Peptone No.3 (BDH); 1.5g $Mg_2SO_4.7H_2O$; 1.5g $K_2HPO_4.3H_2O$ L^{-1}), Limited-pantothenate King's medium (KBp. 10g glycerol; 10g vitamin-free casein hydrolysate acid (BDH); 1.5g $Mg_2SO_4.7H_2O$; 1.5g $K_2HPO_4.3H_2O$; 24μg pantothenic acid L^{-1}), Minimal M9 medium (M9. 6g Na_2HPO_4; 3g KH_2PO_4; 0.5g NaCl; 1g NH_4Cl L^{-1}; 2mM $MgSO_4$; 0.1mM $CaCl_2$; 0.4% glucose), or SOC (20g tryptone; 5g yeast extract (BDH); 0.5g NaCl L^{-1}; 2.5mM KCl; 20mM glucose; 10mM MgCl; pH 7.0), with agar added to 1.5% (w/v) where appropriate. Overnight cultures where gown shaking in 25ml Sterilin tubes containing 5ml of medium. Microcosms consisted of 25ml glass tubes containing 6ml of medium, supplemented with 0.0024% pantothenic acid where necessary, grown in a static incubator. *E. coli* was grown at 37°C, *P. fluorescens* at 28°C. Dilutions were performed in Ringers solution (2.25g

NaCl; 0.015g KCl; 0.12g CaCl; 0.05g $NaHCO_3$ L^{-1}). Bacterial strains were stored indefinitely at -80°C in 20% glycerol.

2.1.4 Antibiotics, enzymes, and reagents

Antibiotics (Melford Laboratories) were filter-sterilised and used at the following concentrations: Tetracycline 12.5μg ml^{-1} (in LB), 37.5μg ml^{-1} (in KB and PAF); ampicillin 100μg ml^{-1}; kanamycin 25μg ml^{-1} (*E. coli*), 75μg ml^{-1} (*P. fluorescens*). Restriction endonucleases and T4 DNA ligase (New England Biolabs); *Taq* polymerase and *Pfu* polymerase (Qiagen); and BigDye sequencing mix (Genosys) were used in accordance with the manufacturers' instructions. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal; 45μg ml^{-1}; Sigma) and o-Nitrophenyl--β-D-galactopyranoside (ONPG; 4mg ml^{-1}; Sigma) were used as chromogenic indicators of β-galactosidase activity. Congo Red (Sigma) was used as a qualitative and quantitative stain for cellulose at the appropriate concentrations.

2.1.5 Oligonucleotides

Oligonucleotides for PCR and sequencing were synthesised by MWG-Biotech and stored at 100pmol μl^{-1} (PCR) or 1pmol μl^{-1} (sequencing) in MilliQ water. The following primers were used:

Primer Name	**Primer Sequence (5'-3')**
*Wsp*R-S1	GCACATGAATGATTTACAGATCG
*Wsp*R-S2	CCAACGATTATTTGGTCAAGC
*Wsp*R-S3	ATTCGTGAGGCCAGCAGTC
*Wsp*R-S4	GAACCATTAGTGGCCGTGAC
*Wsp*R-S5	CTGGCCTCACGAATGGTC

*Wsp*R-S6	CTCTTGATCAGCGGGTCTTC
*Wsp*R-for	CGCGGATCCGAGGAGACTGCACATGAATGATTTAC
*Wsp*R-rev	CCGGAATTCTCGGCTCTAAGGCCGCCGCG
*Wsp*R-DN-for	TTATCCTGCAGAATCTGGTGATG
*Wsp*R-DN-rev	CATCACCAGATTCTGCAGGATAA
*Wsp*R-forA	CGCAAGCTTTCAGGTGACTGCACATGAATGATTTAC
*Wsp*R-midrevA	GCAGATAGTCCTGGCTGACCCGCAACGCGCGA
PleD-midforA	GGTCAGCCAGGACTATCTGCGCAACAATCT
PleD-revA	CGAATTCAGTCAGGCGGCCTTGCCGACCAC
*Wsp*R-midrevB	TCAGGTGGTCGGAGTTCATCAGCCGTTGCAGC
PleD-midforB	GATGAACTCCGACCAGCTGACCGGCCTGCACAAT
*Wsp*R-midforC	GCGGATCCGGTACCCTAGGTGTGACGTGATGTTGGA
*Wsp*R-midrevC	CGGAATTCGGTACCCTAGGAGCGCGAGTGATAGCG
WspR-midrevD	CGGAATTCGGTACCACAACAGGGTCATGTAGGAGCG

2.1.6 Electrophoresis materials

Agarose gels were made to 0.8%, unless otherwise stated, and were run in 1xTAE (40mM Tris pH8; 1.142ml glacial acetic acid L^{-1}; 1mM EDTA) or 0.5xTBE (90mM Tris-HCl pH8; 0.55% boric acid; 2mM EDTA). Gels contained 10μg ml^{-1} ethidium bromide for UV visualisation of DNA. Samples were loaded in Sample Loading Buffer III (Sambrook *et al.* (1989) 0.045% bromophenol blue; 0.04% xylene cyanol FF; 5% glycerol, in water). Polyacrylamide gels were made to various percentages with 37.5:1 acrylamide:bisacrylamide. Stacking gels were made to 5% using 250μl 0.5M Tris pH6.8, 20μl 5% sodium thiosulphate, 10μl 10% SDS, 7μl 10% ammonium persulphate, 1.4μl TEMED ml^{-1}; resolving gels were made to the required percentage using 250μl 1.5M Tris pH8.8, 20μl 5% sodium thiosulphate, 10μl 10% SDS, 7μl 10% ammonium persulphate, 1.4μl TEMED ml^{-1}. Gels were run in PAGE

Running Buffer (25mM Tris; 192 mM glycine; 0.15 SDS; pH8.3) and samples were loaded in PAGE Sample Buffer (4% SDS; 125mM, Tris pH6.8; 20% glycerol; 0.004% bromophenol blue; 10% mercaptoethanol). Pre-cast PAGE gels were purchased from Biorad. Immobilised pH gradient (IPG) strips (7cm, pH3-10 NL) were purchased from Amersham.

2.2 Methods

2.2.1 Conjugation

2.2.1.1 Biparental mating

Overnight cultures of donor (*E. coli* S17-1 *λpir*) and recipient (*P. fluorescens* or *C. crescentus*) strains were washed and resuspended in sterile LB to remove antibiotics. Recipient cells were incubated at 45°C for 15 minutes to turn off host restriction-modification system, while donor cells were incubated for the same length of time at 37°C. 700μl recipient and 300μl donor were mixed in an Eppendorf tube, spun down in a microcentrifuge at 13,000rpm for 1 minute, and resuspended in 50μl LB. This was then pipetted onto a LB agar plate on which it was left as a droplet, and the plate was incubated at 28°C for 4 hours. The dried droplet was removed with a sterile toothpick and resuspended in 1ml Ringers Solution. When integration of the plasmid was not sought, this suspension was diluted to 10^{-3} before plating. 100μl of the suspension (when looking for integration) or the dilution (in all other cases) was plated onto LB or PAF plates, containing ampicillin to select against *E. coli*, and the appropriate antibiotic to select against non-transconjugants. Colonies were taken from these plates after growth for two days at 28°C.

2.2.1.2 Triparental mating

When the donor strain was DH5α, which does not contain *tra* and *mob* genes for conjugation, a helper plasmid possessing these genes was required. The *E. coli* strain containing the helper plasmid, pRK2013, was grown and washed in the same way as the donor strain in the bipartite mating, and was added to the donor strain in a 1:3 ratio before incubation at 37°C. Subsequently the procedure was identical to that in the bipartite mating.

2.2.2 Electrophoresis

2.2.2.1 Agarose gel electrophoresis

Agarose gel electrophoresis was performed in a Mini Sub DNA Cell (Biorad) in accordance with the manufacturer's instructions. Gels were typically run at 80V until sufficient resolution was achieved for the bands of interest. Unless otherwise stated, the 100bp ladder (NEB) was used as a standard. Analytic gels were run in TBE, while preparative gels were run in TAE. DNA was visualised using a UV Transilluminator and Gel Documentation System (UVP).

2.2.2.2 Polyacrylamide gel electrophoresis (PAGE)

2.2.2.2.1 1D SDS-PAGE

One dimensional PAGE was performed in a Mini Protean II Cell (Biorad) in accordance with the manufacturer's instructions. Samples were prepared as follows (after Xu and Gross, 1988): 50ml of overnight culture was resuspended in distilled water and then repelleted; the pellet was resuspended in 500µl sample buffer (0.125M Tris-HCl, pH6.8; 4% SDS; 20% glycerol; 1.4M β-mercaptoethanol) and then heated at 37°C for 1 hour; after centrifugation at 13,000rpm for 10 minutes, 20µl of sample was loaded onto the gel. Gels were run at 21mA through the stacking gel and then at 25mA until the dye-front reached the bottom. Unless otherwise stated Broad Range Protein Markers (NEB) were used as a standard. Proteins were fixed to the gel by incubation in 12.5% TCA for one hour. Proteins were visualised by staining overnight in Biosafe Coomassie Stain (Biorad) and then the gel was washed in 40% methanol before storage in MilliQ water.

2.2.2.2.2 2D SDS-PAGE

Samples were prepared for two dimensional PAGE as follows: 50ml of overnight culture was centrifuged at 4,000rpm and 4°C for 30 minutes and then washed 3 times in low salt wash buffer (3mM KCl; 1.5mM KH_2PO_4; 68mM NaCl; 9mM NaH_2PO_4). The pellet was

resuspended in 600µl lysis buffer (10mM Tris-HCl, pH8; 1.5mM $MgCl_2$; 10mM KCl; 0.5mM DTT, 0.1% SDS) and stored at -20°C. Protein content was assayed using the Biorad Protein Assay (after Bradford, 1974) in conjunction with a 20 Genesys spectrophotometer (Spectronic Instruments) both according to the manufacturer's instructions. Sample was added to IPG buffer (8M urea; 2% CHAPS; 2% IPG buffer (Amersham); 0.3% DTT, 1µl bromophenol blue) to result in a total final volume of 125µl and a total final protein content of 250µg. This was centrifuged (13,000rpm; 1 minute) and the supernatant was added, with an IPG strip, to a re-swelling tray (Amersham), covered with 3ml mineral oil, and left overnight. Strips were removed from the re-swelling tray, washed, dried and run in an Immobiline Drystrip Tray (Amersham) at 3000V for 16 hours according to the manufacturer's instructions. Strips were washed and equilibrated in 2nd dimension buffer (0.05M Tris, pH6.8; 6M urea; 4% SDS; 20% glycerol; 10% β-mercaptoethanol; trace bromophenol blue) for 15 minutes. The strip was then loaded onto a polyacrylamide gel with appropriate well and run as for 1D SDS PAGE.

2.2.3 DNA purification and cloning

2.2.3.1 Digestion, purification and ligation

Cloning was carried out in accordance with standard molecular biology techniques (Sambrook *et al.* 1989). Specifically, fragments were either purified from agarose gels using a QIAquick Gel Extraction Kit (Qiagen), or purified from a PCR reaction using a QIAquick PCR Purification Kit (Qiagen). Plasmid DNA was prepared using a Spin Miniprep Kit (Qiagen). After digestion of insert and vector with appropriate restriction endonucleases, and dephosphorylation of vector DNA where necessary, enzymes were heat deactivated at the temperature recommended by the manufacturer. When heat deactivation was not possible, the DNA was ethanol precipitated as follows: 2 volumes of ethanol (AnalR grade) and 1/5 volume 1.5M sodium acetate were added to the reaction and incubated at -20°C for 30 minutes; after centrifugation at 13,000rpm and 4°C for 15 minutes the pellet was washed in

70% ethanol and then air-dried; the pellet was then resuspended in appropriate buffer. Ligations were carried out overnight at 16°C. Before electroporation the ligation mixture was ethanol precipitated, as above, to remove salts.

2.2.3.2 Calcium transformation

Competent *E. coli* cells were produced from cultures grown to mid-log phase in LB. After centrifugation at 10,000rpm for 10 minutes cells were resuspended in 0.1 volume ice-cold 50mM $CaCl_2$. After incubating for 30 minutes on ice this procedure was repeated using 0.05 volumes $CaCl_2$. Glycerol was added to 15% and aliquots of cells were stored at -80°C. For transformation competent cells were thawed on ice before mixing with ligation mixture in a 4:1 volume ratio. After 30 minutes incubation on ice cells were heat-shocked at 42°C for 3 minutes. 9 volumes of LB were then added and the cells were incubated at 37°C for one hour before plating onto selective media. Intact vector DNA was used as a control to estimate transformation efficiency.

2.2.3.3 Electroporation

Electrocompetent cells were produced from cultures grown to mid-log phase in LB. After centrifugation at 5,000rpm at 4°C for 30 minutes, cells were washed in, progressively, 1, 0.5 and 0.25 volumes of ice-cold MilliQ water. The ice-cold pellet was then resuspended in 0.1 volume 10% glycerol, centrifuged as before, and resuspended in 0.05 volumes 10% glycerol. Aliquots of this suspension were stored at -80°C. For electroporation 5µl DNA was added to 20µl competent cells and the mixture was dispensed into a chilled 0.1cm electroporation cuvette (Biorad). An electric field was applied to the cuvette (1.8V, 200Ω, 25µFD) using a Gene Pulser and Pulse Controller (both Biorad). The cells were then transferred to an Eppendorf tube containing 1ml SOC and incubated at 37°C for one hour, before plating onto selective media.

2.2.4 Polymerase Chain Reaction (PCR)

2.2.4.1 Standard protocol

The polymerase chain reaction was carried out using a PTC-100 programmable thermal cycler (MJ Research). Reaction mixture consisted of: 10µl 10x buffer, 5µl 4mM dNTP mix, 3µl 25mM $MgCl_2$, 0.75µl polymerase (all Qiagen); 1µl each primer; 1µl template DNA or 3µl overnight cell culture; made up to 100µl with MilliQ water. After an initial denaturation step at 94°C for 10 minutes amplification was performed by 30 cycles of denaturation at 94°C, annealing the appropriate temperature and extension at 72°C. Relative times for the steps were varied depending on the DNA for amplification, though typical values were 40 seconds denaturation, 30 seconds annealing, and 2 minutes extension. A final extension step at 72°C for 10 minutes was performed before the sample was cooled indefinitely at 4°C.

2.2.4.2 Strand Overlap Extension (SOEing)

Strand overlap extension was used for creating gene fusions and for site-directed mutagenesis, and is illustrated in Figure 2.1. PCR products of the two fragments for fusion, or the DNA either side of the mutation, were created using the standard protocol above, with primers that complemented for 20bp at the site for joining (and containing the desired mutation in the case of site-directed mutagenesis). These fragments were gel purified using a QIAquick Gel Extraction Kit (Qiagen). For the annealing step the following mixture was used: 5µl each purified DNA fragment; 4µl 10x buffer, 2µl 4mM dNTP mix, 1µl 25mM $MgCl_2$, 0.5µl polymerase (all Qiagen); made up to 40µl with MilliQ water. This was incubated at 94°C for 10 minutes, before 8 cycles of the following: 94°C for 2 minutes, 53°C for 1 minute and 72°C for 4 minutes; and then 10 minutes further extension at 72°C. For the amplification step the following mixture was used: 2µl 10x buffer, 1µl 4mM dNTP mix, 1µl 25mM $MgCl_2$, 0.5µl polymerase (all Qiagen); 1µl each external primer; 30µl annealing reaction. This was cylced

for 20 cycles of the following: 94°C for 1 minute, 55°C for 1.5 minutes, 72°C for 2.5 minutes; followed by 10 minutes further extension at 72°C and indefinite cooling at 4°C.

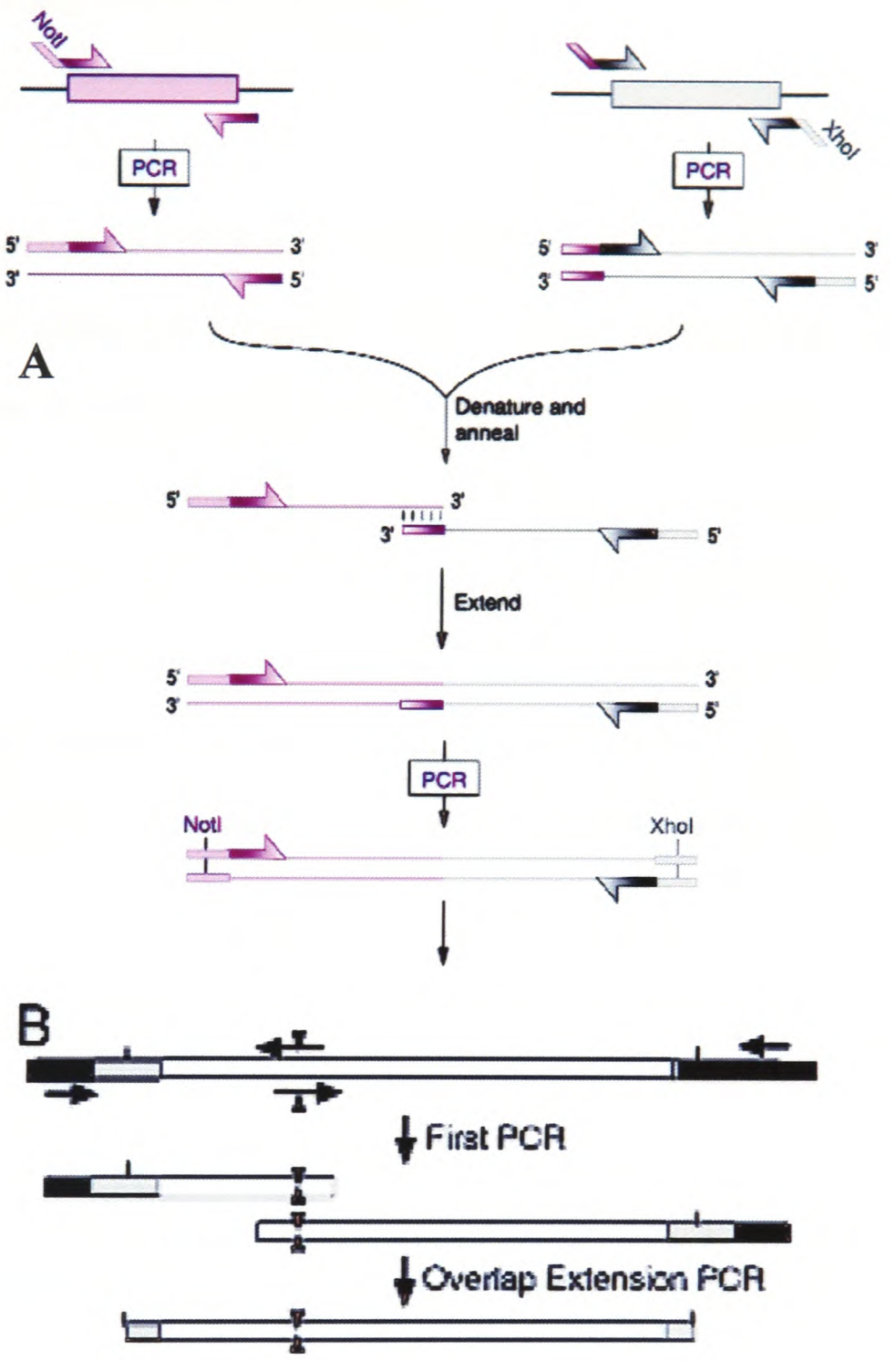


Figure 2.1. Strand Overlap Extension. Schematic representation of the SOEing technique. A: An example of the use of the technique for the creation of gene fusions, shown here using the enzymes *Not*I and *Xho*I. The forward primer used in the second gene is complementary to the reverse primer in the first gene allowing annealing and subsequent extension. B: An example of the use of the technique for site-directed mutagenesis. The internal primers contain the mismatch that gives the desired mutation. The two products are then annealed and extended as above.

2.2.5 Congo Red staining for cellulose

2.2.5.1 Qualitative plate assay

As a qualitative measure of cellulose production, bacteria were grown on agar plates containing Congo Red. As Congo Red is precipitated by high salt concentrations, LBns media was used (except when salt was titrated into the media when assessing salt-dependent phenotypes.) Congo Red was added to 0.0002%. 5µl overnight LB culture was dropped onto

the plate without spreading and left still until dried. The plate was then incubated at 28°C for 24 hours, before phenotypes were scored. Strains that did not produce cellulose remained yellow, while those that did bound Congo Red and appeared red.

2.2.5.2 Quantitative assay

To achieve a quantitative measure of cellulose production, Congo Red binding was measured in liquid culture. 1ml overnight LB culture was stained for 1 hour in 0.001% Congo Red at 37°C (*P. fluorescens* does not grow at this temperature). The culture was washed twice in 0.5ml LB before resuspension in 90% ethanol. After 1 hour shaking incubation at 37°C this was centrifuged at 13,000rpm for 1 minute. The A_{500} of the supernatant was measured using a 20 Genesys spectrophotometer (Spectronic Instruments) against a 90% ethanol blank, and the absorbance was compared with appropriate Congo Red controls. The lower the absorbance, the greater the amount of Congo Red that was bound by the cells, and hence the greater the amount of cellulose produced.

2.2.6 Microcosms

2.2.6.1 Niche-preference assay

Microcosms, containing the appropriate antibiotic, were inoculated with 6µl overnight LB culture, and incubated statically at 28°C for 16 hours with lids loosely fitted. Niche preference was scored qualitatively by looking for the presence or absence of a biofilm, and the presence or absence of observable cell densities growing below the air-broth interface. Each microcosm was recorded as strong (++), medium (+), or negligible (–) for both surface and sub-surface growth.

2.2.6.2 Competitions

2.2.6.2.1 Standard protocol

For competition assays one of the two genotypes to be competed was a pantothenate mutant, a trait that is selectively neutral in media supplemented with pantothenate, but can be identified on KBp plates by the dramatically reduced colony size. The two genotypes to be competed were grown overnight in LB and a 1:1 volume ratio mixture was made. 60µl of this mixture was used to inoculate the KBp microcosm, while a 10^{-6} dilution of the mixture was plated onto KBp (plates and microcosms both contained the appropriate antibiotics). The microcosms were incubated at 28°C for the desired number of days, either shaking or static as required. Microcosms were harvested by vortexing for one minute, followed by dilution to 10^{-6} and plating on KBp. Selection coefficients were calculated as the ratio of the Malthusian parameters of the two genotypes: $\ln[100(a_t/a_0)/((a_t+b_t)/(a_0+b_0))] / \ln[100(b_t/b_0)/((a_t+b_t)/(a_0+b_0))]$ assuming a 100-fold increase in population size (where a_0 and a_t = proportion of genotype a in the population at times 0 and t respectively, and b_0 and b_t = proportion of genotype b in the population at times 0 and t respectively). All competitions were performed in triplicate.

2.2.6.2.2 Invasion from rare

To assess the ability of genotypes to invade populations from rare, often as a result of negative frequency-dependent selection, microcosms were set up with 100:1 volume ratio of the two genotypes. Due to the large difference in cell density between the two genotypes serial dilutions had to be plated out to acquire adequate data. To allow time for sufficient selection 6µl of the mixture was used to inoculate the microcosms, rather than the 60µl in the standard protocol, and a 1000-fold increase in the population size was assumed in the calculations. Otherwise the procedure was as for the standard protocol above.

2.2.7 Miller assay

β-galactosidase assays were performed in accordance with the standard protocol (Miller, 1972). 100µl overnight culture or solubilised plate colony was added to 800µl Z buffer

(60mM $NaHPO_4$, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$, 50mM β-mercaptoethanol) and 10µl chloroform and vortexed. After keeping on ice until required the solution was incubated at 28°C for 2 minutes. To start the reaction 160µl 4mg/ml ONPG was added before incubation at 28°C for a further 10 minutes. The reaction was stopped using 400µl 1M Na_2CO_3. After centrifugation at 13,000rpm for one minute a 20 Genesys spectrophotometer (Spectronic Instruments) was used to measure A_{420} and A_{550}, as well as A_{600} of the original cell culture (using appropriate blanks). β-galactosidase activity was calculated using the following formula: $8000(A_{420} - 1.75\ A_{550})/t\ A_{600}$ (t: time of reaction in minutes, typically 10). All assays were performed in triplicate.

2.2.8 DNA sequencing

Samples were prepared for sequence analysis using the following method: DNA was purified by Spin Miniprep Kit (Qiagen) for plasmids, or QIAquick PCR Purification Kit (Qiagen) for PCR products and 0.1µg was added to 4µl BigDye Reaction Mix, 0.5µl primer and MilliQ water to 10µl; this was heated in a PTC-100 programmable thermal cycler (MJ Research) for 25 cylces of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes; the reaction was ethanol precipitated as above and left as a dry pellet. This sample was submitted to the Plant Sciences Departmental Sequencing Facility for analysis on an ABI Prism 310 Sequencer. The resultant electrophoretograms were analysed using ABI Sequence Navigator software.

2.2.9 Microscopic analysis of *C. crescentus* stalk cells

Strains were analysed by phase-contrast light microscopy, using an Olympus AX-70 microscope. The ratio of stalk cells to non-stalk cells was calculated from the analysis of 200 pre-divisional cells taken from log-phase culture and immobilised on microscope slides with poly-L-lysine (Sigma). The ratio was expressed as a percentage and three replicates were

performed for each strain. Comparsion was then made with the wild-type. This work was kindly carried out by Phillip Aldridge at The University of Basel, Switzerland.

2.2.10 Motility assay

Overnight cultures of cells were grown up in LB to an OD_{600} of 0.6. Cultures were adjusted to the same absorbance where necessary. A pin was used to stab inoculate a plate with the culture; the plates were 0.1 strength KB and 0.35% agar to force semi-starvation and hence induce motility. After 20 hours the diameter of the swarm was measured at two perpendicular points and the mean taken. For each strain five replicates were performed.

2.2.11 Attachment assay

Quantitative attachment assays were performed using a modification of the method of O'Toole *et al.* (1999). KB microcosms, containing the relevant antibiotics, were inoculated from 6µl of overnight culture. After 24 hours of growth the microcosm was emptied and washed vigorously with dH_2O. 10ml 0.05% Crystal Violet stain was added to the tube and left for two minutes. The stain was then eluted with 5ml ethanol shaken for 30 minutes. The OD_{570} of the ethanol was then determined as a measure of staining.

2.2.12 Statistical analysis

Where multiple data points were taken (Miller assays and competition experiments) means and standard errors were calculated, the latter being the error bars on all graphs. ANOVA was used to detect significant effects of treatments, with the one-way and two-way cases being used as appropriate; these are displayed in the appendix. Significant differences between individual samples were tested using a two-tailed t test adjusted for the Bonferroni correction as necessary.

Chapter 3

Analysis of *wspR* Mutant Alleles

Previously, the response regulator WspR was identified as necessary for the evolved LSWS phenotype. In this chapter seven point mutants of the *wspR* gene are sequenced and then overexpressed in the ancestral SM and evolved LSWS genotypes. The phenotypic effects of the different alleles are assessed with a range of different assays, and inferences are made about the structure-function relationship within the protein sequence and about the mechanism by which WspR affects cellulose overproduction.

Aims:

1. To sequence the *wspR* alleles
2. To establish a structure-function correlation
3. To identify the factors that input into WspR
4. To assess likely mechanisms of WspR output
5. To propose a model of WspR function

3.1 Introduction

As discussed in Chapter 1, a transposon mutagenesis screen of the evolved biofilm forming genotype, The Large Spreading Wrinkly Spreader (LSWS), identified the response regulator-encoding gene, *wspR*, as necessary for the surface-colonising phenotype (Spiers *et al.* 2002). This raises both the mechanistic question of what role *wspR* plays in the overproduction of cellulose, and the evolutionary question of where the selected mutation is located and its relationship to *wspR*. In order to understand fully the transition between the ancestral SM and LSWS both these questions must eventually be answered.

The first approach, taken in this chapter, was designed to give further clues about the answers to both questions in order to be able to direct and refine future enquiries. This approach was the analysis of random point mutations in *wspR* and their phenotypic effects. The rationale was to examine the structure-function relationship within the sequence of *wspR* in respect of the first question, and to look at candidate mutations for adaptive change in respect of the second. As discussed below, the mutations studied were generated at random, and are artefacts rather than alleles found in naturally evolved genotypes. Such an approach of random mutagenesis, if carried out to saturation, could potentially identify all the residues in WspR that are necessary for function in a non-redundant way, in the same way that the initial mutant screen of LSWS identified the genes that are necessary for the phenotype. The next chapter uses a more directed approach that involves making specific mutations to the *wspR* sequence, more analogous to making targeted gene-knockouts at the whole-genome level.

The origin of the point mutations analysed remains unclear, though the procedure is described by Kahn (1998). Briefly, *wspR* was cloned independently from PCR products into pVSP61, a *Pseudomonas* expression vector. The PCR products were amplified from various genotypes of smooth, wrinkly and fuzzy morphology, isolated from independent microcosms. The rationale behind this was that changes in the sequence of *wspR* might be responsible for the

observed variation in phenotype. However, it is no longer thought that the sequences of the cloned alleles correspond to the sequences of the alleles found in the chromosomes of the genotypes from which they were taken. Instead it is thought that the mutations arose during the cloning procedure and that all the chromosomes contain a wild-type copy of *wspR*. The reasons behind this will be discussed more fully below.

The *E. coli wspR* clones were conjugated into the ancestral SM genotype and found to result in a range of phenotypes, both SM and WS. Likewise the clones were conjugated into LSWS and a similar range of phenotypes was observed. These phenotypic effects were found to be stable through several rounds of conjugation between the two species as illustrated in Figure 3.1.

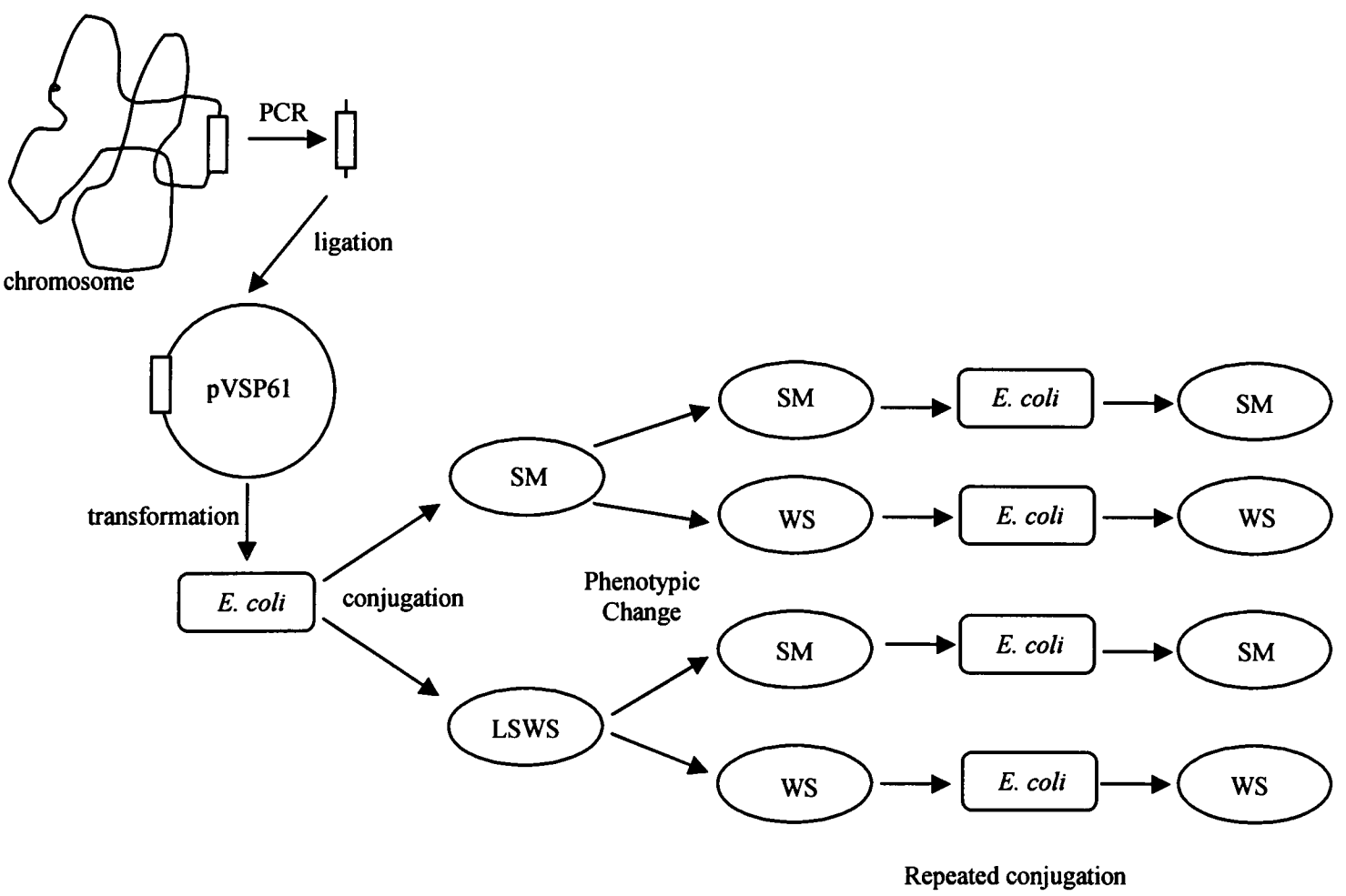


Figure 3.1. The Cloning of *wspR* Alleles. This procedure was performed independently for SM, WS and FS genotypes. *WspR* was amplified by PCR from the chromosome, cloned into pVSP61, and transformed into *E. coli*. The recombinant plasmids in the transformants were transferred *en masse* by conjugation into SM and LSWS genotypes. In each case a range of phenotypes were found amongst the colonies on the conjugation plate, some of which were different form the parental phenotype. Representative colonies were selected and transferred repeatedly between *E. coli* and the relevant *P. fluorescens* genotype and the phenotypes were found to be stable. In this diagram *E. coli* strains are represented by rectangles, and *P. fluorescens* strains by ellipses.

A representative sample of these *E. coli* strains was chosen for use in this study. Sequencing of three alleles was performed by Kahn (1998), but these are confirmed in this chapter at the same time as sequencing the remaining alleles.

In this chapter these *wspR* clones are expressed in different SBW25 genotypes and the phenotypic effects characterised in various biologically relevant ways. The *wsp*R alleles are also sequenced so that a correlation can be established between structure and function.

The alleles used are *wspR5*, *wspR9*, *wspR12*, *wspR13*, *wspR14*, and *wspR19* kindly made available by Sophie Kahn (in the form of the *E. coli* pVSP61 clones produced in Figure 3.1), and *wspR1* produced during this study.

3.2 Results

3.2.1 Bioinformatic analysis of *wspR*

Kahn (1998) reported that the sequence of the cloned *wspR* allele 12 was the same as the copy found in the chromosomes of both ancestral SM and the evolved LSWS. As this gene will form the basis of this study this result was checked in order to gain a definitive wild-type sequence for *wspR*. PCR products amplified from the chromosomes of the two genotypes and also from the plasmid containing *wspR*12 were sequenced in the manner that is described fully in Section 3.2.3. This procedure confirmed the sequence reported by Kahn (1998), and is used for subsequent bioinformatic analyses, most often translated into amino acid sequence.

Preliminary tBLASTn searches [which use amino acid sequence (Altschul *et al.*, 1990)], with the WspR protein sequence, recovered sequences which had significant similarity to either one or other end of the sequence, but rarely to both. This implies that *wspR* has two domains, which are probably of different evolutionary origin as they are found separately more often than together. At the very least they must have been shuffled to create the homologues even if the *wspR* configuration is ancestral. This seems unlikely given the large numbers of alternative configurations, many of which are found many times in sequence databases. WspR homologues across the entire sequence length are much less common, as discussed in Chapter 1.

In order to define these two domains the ten sequences of highest similarity to each terminus were selected for PILEUP (Genetics Computing Group) analyses. For each set of ten a separate PILEUP was obtained and the results are shown in Figures 3.2 and 3.3.

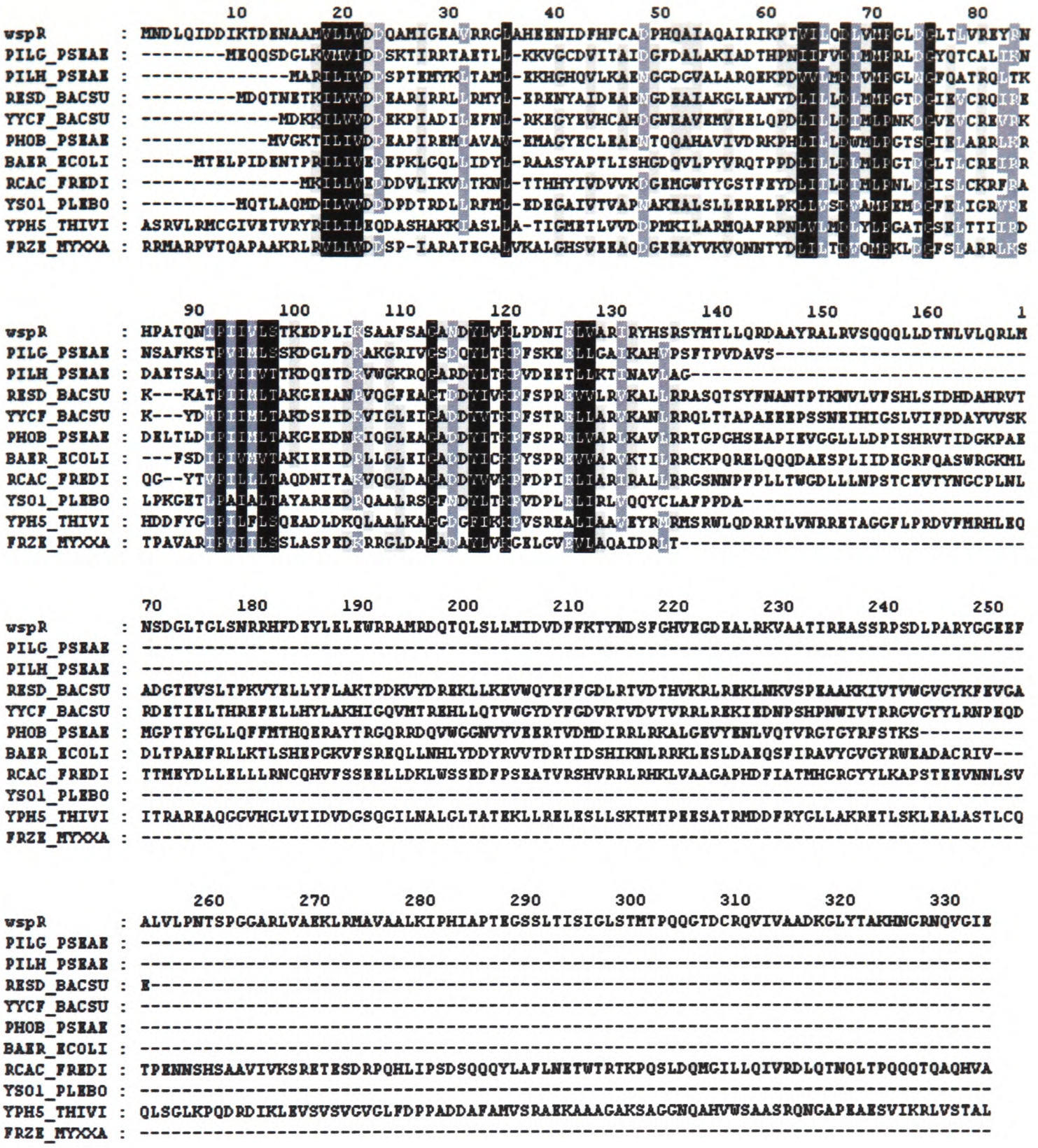


Figure 3.2. WspR N terminus. A multiple sequence PILEUP is shown of the WspR protein sequence and the 10 translated sequences that scored most highly to the N terminus when a tBLASTn search of all databases was performed. Alignments are shaded black for 100% conservation, dark grey for 80% and light grey for 60%. Where sequences were longer than WspR at either end the extra residues are not shown. The region of significant conservation ends at residue 130 and thereon there is little conservation. Among the conserved residues in the N terminus is the aspartate at residue 67 which is a putative phosphorylation site. The ten protein sequences are from the genes *pilG*, *pilH* and *phoB* from *P. aeruginosa*; *resD* and *yycF* from *B. subtilis*; *baeR* from *E. coli*; *rcaC* from *Fremyella diplosiphon*; *ys01* from *Plectonema boryanum*; *yph5* from *Thiocystis violacea*; and *frzE* from *M. xanthus*.

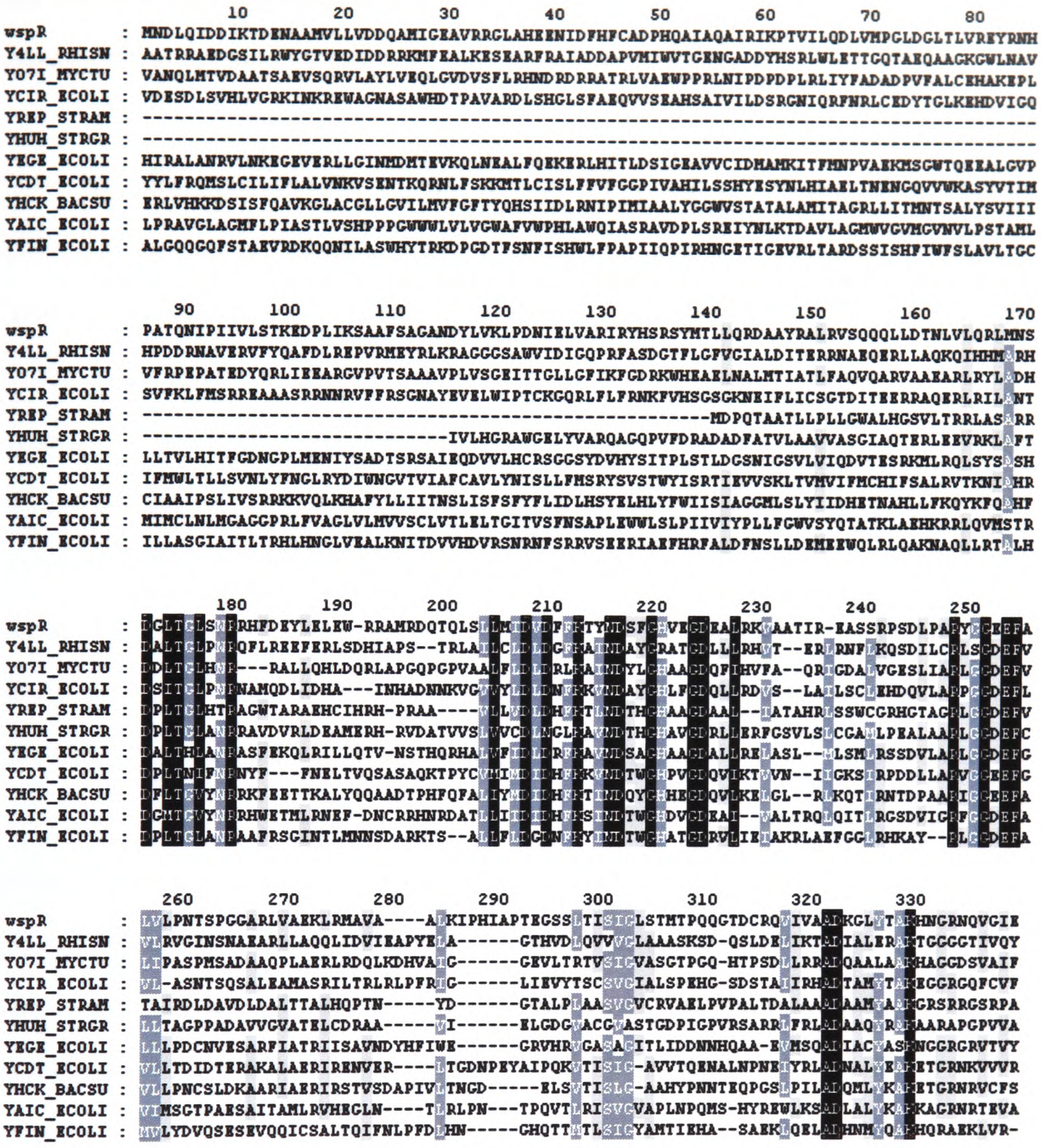


Figure 3.3. WspR C terminus. A multiple sequence PILEUP is shown of the WspR protein sequence and the 10 translated sequences that scored most highly to the C terminus when a tBLASTn search of all databases was performed. Alignments are shaded black for 100% conservation, dark grey for 80% and light grey for 60%. Where sequences were longer than WspR at either end the extra residues are not shown. There is little conservation in the N-terminal region, while the notable conservation starts at residue 164. Included among the highly conserved sections of the C terminus is the GGDEF motif. The ten protein sequences are from the genes *y4lL* from the *Rhizobium* species plasmid pNGR234a; *yo7I* from *Mycobacterium tuberculosis*; *yciR*, *yegE*, *ycdT*, *yaiC* and *yfiN* from *E. coli*; *yreP* from the *Streptomyces ambofaciens* plasmid pSAM2; *yhuH* from *Streptomyces griseus*; and *yhcK* from *B. subtilis.*

It can be clearly seen that the homology for the N terminal sequences ends at amino acid 130, while the homology for the C terminal sequences begins at amino acid 164. Hence the protein

can be said to have two domains separated by a linker sequence. This latter, which runs from residues 131 to 164 has little similarity to any sequence in the database.

Following the example of Pei and Grishin (2001), who searched for homologues of GGDEF proteins at the level of three-dimensional domain structure, tertiary structural homologues of WspR were identified. This was performed using the hybrid fold recognition method (Fischer, 2000) via the BIOINBGU server (http://www.cs.bgu.ac.il/~bioinbgu/). When the full length sequence of WspR was used the two domains showing greatest similarity (with respective consensus scores of 50.9% and 31.9%) were the phosphorylated aspartate receiver domain of Spo0a from *Bacillus stearothermophilus*, and the receiver domain of PhoB from *E. coli*. Other significant matches included the receiver domain of FixJ from *Rhizobium meliloti*, CheY and CheB from *E. coli*, and the receiver domain of NtrC from *S. typhimurium*. These are all expected homologues of the N-terminal domain, and when that domain alone was submitted to the server (using the sequence of *wspR*NtermB as discussed in Chapter 4) the same homologues were seen, but with proportionally higher consensus scores (87.0% for Spo0a). When the C terminus alone was submitted (using the sequence of *wspR*Cterm as discussed in Chapter 4) the domains retrieved were completely different and the scores were much lower. This is to be expected as no crystal structure for GGDEF domains exists, and this is the reason that Pei and Grishin (2001) performed their study to identify adenylate cyclase as a structural homologue. The highest match to WspR, with a consensus value of 6.5%, was guanylate cylase, while other nucleotide binding proteins also featured in the diverse range of domains retrieved. These results are consistent with those of Pei and Grishin (2001) but the numbers are similarly low and caution should be taken in their interpretation.

3.2.2 Colony morphology and production of *wspR1*

Kahn (1998) reported that the conjugation of pVSP61, a *Pseudomonas* expression vector (Loper and Lindow, 1994), containing the 6 different *wspR* alleles, into SM and LSWS

genotypes caused changes in colony morphology. This result was confirmed, and is discussed more fully below, in the context of niche preference. As shown in Figure 3.1 the origin of the *wspR* alleles remains unclear despite the transparency of the process by which they were produced. It was originally thought that the alleles corresponded to the different versions of *wspR* in the genotypes that they were taken from (SM isolates in the cases of alleles *wspR9* and *wspR19*, WS in *wspR13* and *wspR14*, and FS in *wspR5* and *wspR12*). Work discussed in Chapter 6 suggests that this is unlikely because direct sequencing from independent microcosm isolates shows no variation in the sequence of *wspR*, despite phenotypic differences, while some of the genotypes that would appear to have given different *wspR* sequences are phenotypically similar (Kahn, 1998). Another possibility is a semi-toxic interaction with *E. coli* during the cloning process resulting in only non-functional copies being viable despite making up only a small fraction of the population. A further line of evidence suggesting this is the observation (P.B. Rainey, personal communication) that spontaneous deletions of *wspR* arise in the attempt to clone it into pBluescript, and also that different alleles have different effects on the host growth curve. However, the fact that the wild-type allele is one of those that have been successfully cloned into the expression vector, as well as several other alleles that retain or even enhance function, makes this explanation unlikely. The third possibility is PCR error when amplifying from the original strains. It was thought that the use of *Pfu*, a proof-reading polymerase, made this unlikely, especially at the high frequency observed (5 out of 6 of the original clones were mutant). However, when a single PCR reaction was cloned into pVSP61 and the transformants were conjugated *en masse* into an SM and LSWS genotypes, a range of phenotypes were present on the same plate, with under 80% being of the parental phenotype. If the PCR is error-prone this may be a property of *wspR* itself, though not necessarily a biologically interesting one. This cloning was carried out according to Kahn (1998) using the primers *wspR*-for and *wspR*-rev and the restriction enzymes *Bam*H1 and *Eco*R1 respectively. A new allele, *wspR1,* was isolated from this mass conjugation, having the phenotype of turning a WS morph back into SM. The SM colony in question, that had reverted from the LSWS, was taken and plasmid DNA was extracted and

re-transformed into *E. coli*, from where it could be manipulated and sequenced. It should be noted that the nature of the screening procedure means that conservative mutations will not be picked up because they will not alter the phenotype of the parental strain; this applies to non-synonymous substitutions and those that affect amino acid sequence, but not function.

3.2.3 Sequencing of *wspR* alleles

To establish why Kahn (1998) observed such a range of phenotypes caused by different *wspR* alleles, the six alleles from that study and *wspR*1 were sequenced. In the cases of alleles *wspR9*, *wspR12* and *wspR19* this was to confirm the results of Kahn (1998), while the other alleles were sequenced for the first time. Sequencing was performed on *wspR* PCR products which had been amplified from the pVSP61 clones with primers *wspR*-for and *wspR*-rev using *Taq* polymerase. Though *Taq* has a higher error rate than proof-reading polymerases such as *Pfu*, the nature of the PCR and sequencing reactions is such that a consensus sequence emerges from the latter despite errors in individual products from the former. At any given base the majority of products will be wild-type even if each product is mutant in at least one base. It was, however, important not to use the PCR primers for sequencing as this would have meant that any inappropriate PCR product, which shared the primers, would also be sequenced, thus interfering with the sequencing reaction. Six separate sequencing reactions were performed on each allele: three forward reactions using primers S1, S2 and S3 allowed coverage of the 1002bp gene with good quality sequence up to the overlaps, while the three reverse primers S4, S5, and S6 were used to confirm the sequence in case of ambiguities (the positions of the primers in the *wspR* sequence are shown in Kahn, 1998). Each allele was found to differ by at least one nucleotide from the wild-type sequence. The mutations and their effects on protein sequence are summarised in Table 3.1. As can be seen *wspR12* has the wild-type *wspR* sequence [as obtained from WS4 (Kahn, 1998)] while the other alleles all contain non-synonymous substitutions.

Allele	DNA Position	Mutation	Amino Acid Position	Amino Acid Change
1	617	a → g	206	D → G
5	755	t → c	252	F → S
9	886	g → a	296	G → R
13	826, 959	g → a, a → g	276, 320	A → S, Y → C
14	476	a → g	159	D → G
19	385	c → t	129	R → C

Table 3.1. Positions of *wspR* Mutations
Mutations in DNA sequence with respect to the wild-type (*wspR12*) present in each numbered *wspR* allele and the resultant change in amino acid sequence.

It is notable that all the mutations observed are transitions rather than transversions. This may be relevant with respect to the mechanism responsible for the mutations. On the basis of these results and the bioinformatic study, *wspR19* can be said to have an N-terminal mutation, allele *wspR14* a linker-region mutation, and the other alleles C-terminal mutations. It is also notable that there are two mutations in *wspR13*. Of functional significance is the fact that most of the mutations are at highly conserved sites, as defined by dark shading in Figures 3.2 and 3.3. The exceptions are the first mutation in *wspR13*, which may be coincidental with the second mutation being responsible for the phenotype, and the mutation in *wspR14* for which it was not possible to ascertain conservation as it is in the unconserved linker region.

To follow on from the *in silico* analysis performed on the wild-type *wspR* sequence, the mutants were subjected to the same hybrid fold recognition method. While the domains identified as having high similarity were the same for both wild-type and mutant, the mutations in *wspR19* and *wspR14* altered the relative consensus values in the same manner. For example, both mutations increase the similarity to NarL form *E. coli,* and reduce it to PhoB from the same organism. However, it is unclear whether or not this is merely an artefact of the algorithm. The C-terminal mutants had a range of effects but these are undoubtedly artefactual as the main folds identified were all in the N terminus. Another component of the BIONBGU site output was a PHD secondary structure prediction for each mutant. Three of the C terminal mutations, those in *wspR1*, *wspR5* and *wspR9*, were at positions predicted to

form beta-sheet, while both mutations in *wspR13* were at positions of predicted helix. The mutation in *wspR19* was also at a position of predicted helix. In none of the cases however was the structural prediction altered significantly by the mutation. No structure was predicted for the site of the mutation in *wspR14*, in either that allele or the wild-type. However, interestingly, the presence of the mutation in *wspR19* caused a helix to be predicted at the site of the mutation in *wspR14*, whereas no structure was predicted at this site in the wild-type. Again it is unclear how much is artefact, though it is probably significant that defined structure is predicted at most mutation sites.

3.2.4 Morphology and niche preference changes caused by overexpression of *wspR* alleles

The phenotypic consequences of the point mutations in *wspR* were assessed using plasmid-based overexpression of the alleles (Kahn, 1998). The pVSP61-based expression vectors (Loper and Lindow, 1994) containing each allele were conjugated into SM and LSWS genotypes and the new strains were stored. The initial screen for phenotype was an assessment of colony morphology on KB and LB agar plates, as Rainey and Travisano (1998) had reported a correlation between this and niche colonisation, and Kahn (1998) had reported changes caused by these alleles. To confirm that this correlation still stood for these results a niche preference assay was performed using KB microcosms on all the strains. The results of both these assays are shown in Table 3.2. The control is the pVSP61 vector without insert, to ensure that none of the effects are due to plasmid-carriage or the presence of an antibiotic. The results show that those alleles that revert the phenotype of LSWS to SM (*wspR1*, *wspR5*, *wspR9* and *wspR13*) also abolish mat formation, while the alleles that change SM into WS (*wspR19* and *wspR14*) also alter niche preference accordingly. The wild-type allele, *wspR12*, which will be discussed below, changes SM into WS on LB but not KB, and has an intermediate niche colonisation phenotype, causing SM to grow in both biofilm and planktonic phases.

Allele	Host	LB Morphology	KB Morphology	Mat Formation	Cells in Suspension
Control	SM	S	S	-	++
	LSWS	W	W	++	-
12	SM	W[1]	S[2]	++	+
	LSWS	W[3]	W[3]	++	-
1	SM	S	S	-	++
	LSWS	S	S	-	++
5	SM	S	S	-	++
	LSWS	S	S	-	++
9	SM	S	S	-	++
	LSWS	S	S	-	++
13	SM	S	S	-	++
	LSWS	S	S	-	++
14	SM	W[1]	W	++	+
	LSWS	W[1]	W[3]	++	-
19	SM	W	W[3]	+[4]	-[4]
	LSWS	W	W[3]	++	-

[1] Not identical to LSWS
[2] Slightly dimpled variant of SM
[3] Small variant of LSWS
[4] Low overall cell density

Table 3.2. Colony Morphology and Niche Preference Changes Caused by *wspR* Alleles.
Colony morphologies on LB and KB agar plates, and niche preference in KB microcosms, for SM and LSWS genotypes expressing the *wspR* alleles and the control plasmid. S and W represent smooth and wrinkly colony morphology respectively. The niche preferences are scored so that ++ indicates a substantial mat or suspension cell density, + indicates a slight mat or low suspension cell density, and – indicates no visible mat or cells in suspension.

3.2.5 The conditional phenotype of *wspR12*

Table 3.2 shows that overexpression of *wspR12* (the wild-type allele) in the SM genotype, unlike any of the other allele-genotype combinations, causes different phenotypes on LB and KB plates. To find out the cause of this effect the different ingredients of the two media were removed individually and the phenotypes scored. Also the ingredients of one media were added individually to the other and the phenotypes were similarly scored. In this way it was discovered that the presence of NaCl is sufficient to turn SM(pVSP61-*wspR12*) from being a SM-like strain, into being WS. All other ingredient additions and removals had no observable effect on phenotype. This salt effect was also seen in niche preference assays, though it was less marked as even without salt there was some degree of biofilm formation. To establish

what quantity of NaCl was necessary for the transition, a titration was performed, showing that the WS phenotype occurred at one tenth of the normal salt concentration of LB (1%) and that there was a gradual transition in the phenotype with increased salt. Further tests showed that both KCl and sodium glutamate produced similar effects, though the threshold concentration required for the SM to WS transition was lowest with NaCl. The effect was also seen when NaCl was added to minimal medium, though the wrinkly phenotype is never as pronounced as it is on nutrient-rich medium, presumably due to fewer resources being available for the overproduction of cellulose (cf. Travisano and Rainey, 2000). If the presence of more molecules of wild-type WspR in the ancestral SM allows a salt-induced morphology change, the same might be the case without the extra copies, but at a higher threshold salt concentration. Therefore an attempt was also made to change the ancestral SM into a WS morph by the addition of larger amounts of NaCl. However, no effect was seen at five times the normal LB salt concentration, and at higher concentrations the cells were not viable.

3.2.6 Cellulose production in genotypes overexpressing *wspR* alleles

The wrinkly spreader phenotype is caused by the overproduction of cellulose. It is assumed that the overexpression of *wspR* causes the same phenotype in a similar way, i.e. enhanced activity of the gene products of the *wss* operon. To check that the overexpression of different *wspR* alleles does indeed cause cellulose overproduction, qualitative plate assays were performed using LBns (LB without salt, see Chapter 2) and the results are shown in Figure 3.4.

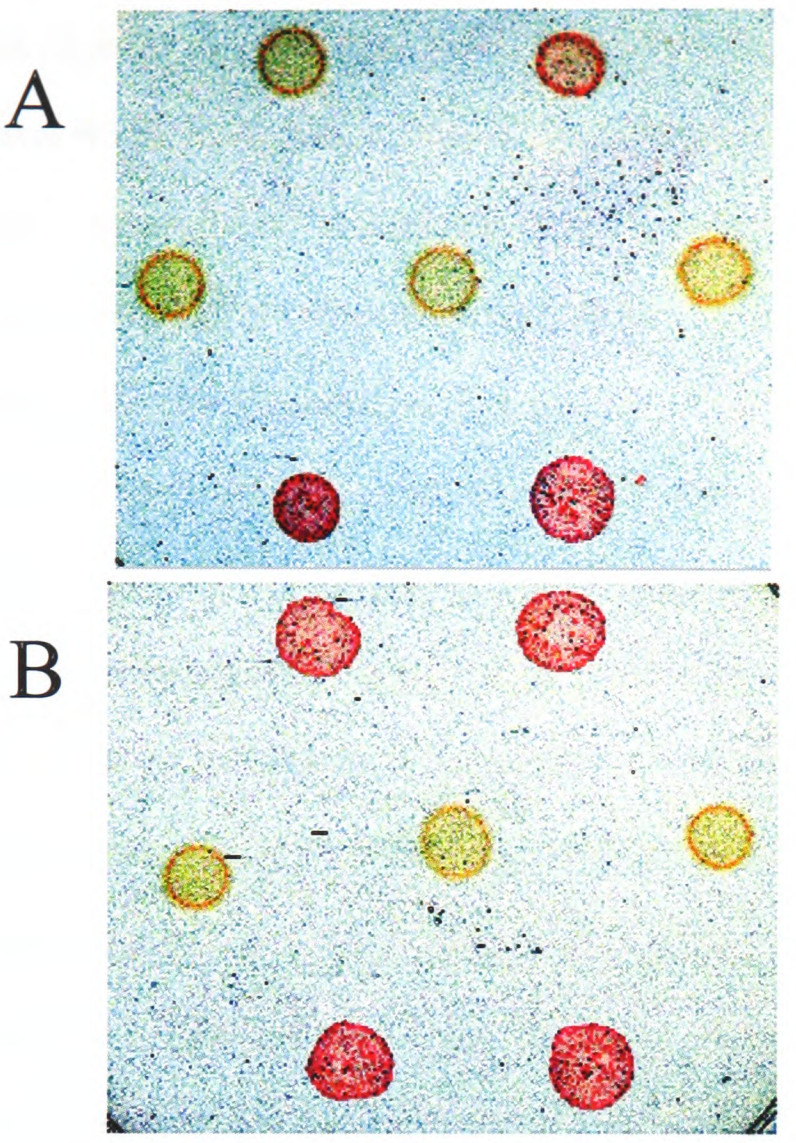


Figure 3.4. Congo Red stains of wspR overexpressing genotypes. Section A shows colonies on a Congo Red plate of SM genotypes overexpressing *wspR* alleles, while Section B shows the same alleles overexpressed in LSWS. In each diagram the alleles are, reading from left to right, top row first, control (i.e. pVSP61 with no *wspR*), *wspR12* (the wild-type allele), *wspR5*, *wspR9* and *wspR13* (the C-terminal mutants), and *wspR14* and *wspR19* (the N-terminal mutants). The genotypes that express cellulose stain red, while those that do not remain yellow with a thin red rim. SM(pVSP61-*wspR12*) shows an intermediate level of staining.

It can be noted that, despite the absence of salt, SM(pVSP61-*wspR12*) has a higher cellulose content than the ancestral SM. NaCl was titrated into the medium and an increase in cellulose was noted, up to the levels seen in the WS control. It was not possible to use LB in the assay as that amount of salt causes the Congo Red to precipitate.

3.2.7 Transcription of the *wss* operon

The transposon mutant WS4 has a SM phenotype (Kahn, 1998), showing that *wspR* is necessary for cellulose overproduction. The conversion of the ancestral SM to a WS phenotype by *wspR*, in the presence of salt, shows that *wspR* is important for cellulose overproduction, but not sufficient without the relevant environmental signals, and possibly

other gene products. In order to know what these other factors might be it was necessary to ascertain at what level WspR is affecting activity of the genes encoding the cellulose synthase complex (*wssB*, *wssC* and *wssE*) and also the rest of the *wss* operon. The first possibility is that it is at the level of transcription, either by direct DNA-binding or through the activities of intermediary proteins. This first possibility is suggested by Aldridge and Jenal (1999) as a possible mechanism for PleD action and, as discussed in Chapter 1, is one of the most common modes of action of response regulators. It has not, however, been demonstrated in any WspR homologues which are thought more likely to function via cyclic di-GMP. To test the possibility that transcriptional activation does occur in *P. fluorescens*, the *wspR* alleles were conjugated into AS14 and AS15, respectively SM and LSWS genotypes with a *lacZ* transcriptional fusion to the end of *wssE*. β-galactosidase activity (representing *wss* operon transcription) was measured using a Miller Assay. In this way it was possible to see whether the different alleles of *wspR* caused the *wss* operon to be expressed at different levels. The fusion to *wssE* is disruptive so that only the first four genes of the operon can be expressed. For this reason any feedback effects on transcription caused by the activity of the downstream gene products would not be seen. A subset of the assays was repeated on cultures grown with and without salt, in both shaken and static broth, as well as on plates, to see if environmental differences caused any compounding effects on *wss* transcription. It might be expected that the *wss* operon would be expressed more highly in a static environment as this is where cellulose biofilm formation occurs, and also in the presence of salt in the light of the conditional phenotype of SM(pVSP61-*wspR12*). The results are shown Figures 3.5 and 3.6. A one-way ANOVA shows a significant effect of genotype on *wss* transcription ($P<0.0001$) in Figure 3.5. In this figure there is a small, but real, difference in expression between genotypes that are phenotypically SM and those that are WS, with the latter being higher. The exception is the case of *wspR19* which has a very large effect on SM and actually reduces *wss* expression in LSWS. Figure 3.6, on the other hand, shows very large differences in *wss* expression in different environments. The presence of salt, counter to expectation, seems to cause lower levels of expression, while, in accordance with expectation, a static environment

generally results in higher levels of expression. A two-way ANOVA of this data shows significant variation in *wss* transcription due to both genotype ($P<0.0001$) and environment ($P=0.0001$), though the interaction between the two just fails to be significant ($P=0.0516$). The genotype effect is expected from the results in Figure 3.5. However, a separate two-way ANOVA of salt presence against environment-genotype combination, shows no significant effect of salt overall ($P=0.253$).

Figure 3.5. Beta-galactosidase activity of SM and LSWS with *lacZ* fusion to *wssE* (genotypes AS14 and AS15 respectively) expressing *wspR* alleles

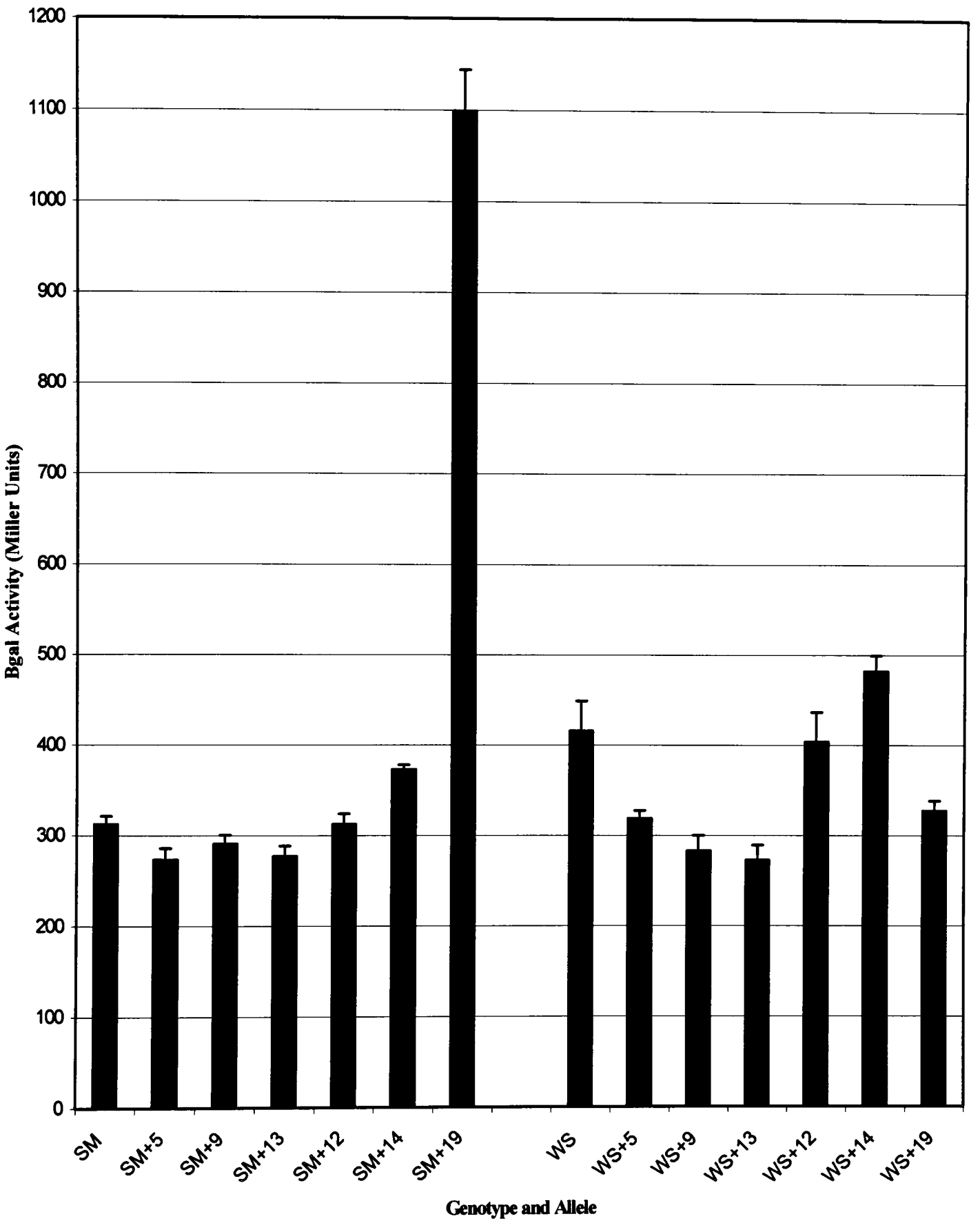


Figure 3.5. *Wss* Transcription. Genotypes containing a *lacZ* fusion to *wssE* and also pVSP61 expressing a *wspR* allele were assayed for β-galactosidase activity after overnight growth. The *lacZ* genotypes, AS14 and AS15, are indicated by SM and WS respectively, the original genotypes to which they correspond. The *wspR* allele present in the pVSP61 vector is also shown, with the controls (the cases of SM and WS) being pVSP61 with no insert. In each data set the plasmid control is shown first, followed by the three C-terminal mutants (*wspR5*, *wspR9* and *wspR13*), the wild-type (*wspR12*), and then the two N-terminal mutants (*wspR14* and *wspR19*). Error bars are the standard error of the three replicates performed in each case.

Figure 3.6. Beta-galactosidase activity of *wssE::lacZ* genotypes AS14, AS14(pVSP61-*wspR12*) and AS15 in three different conditions on two different media

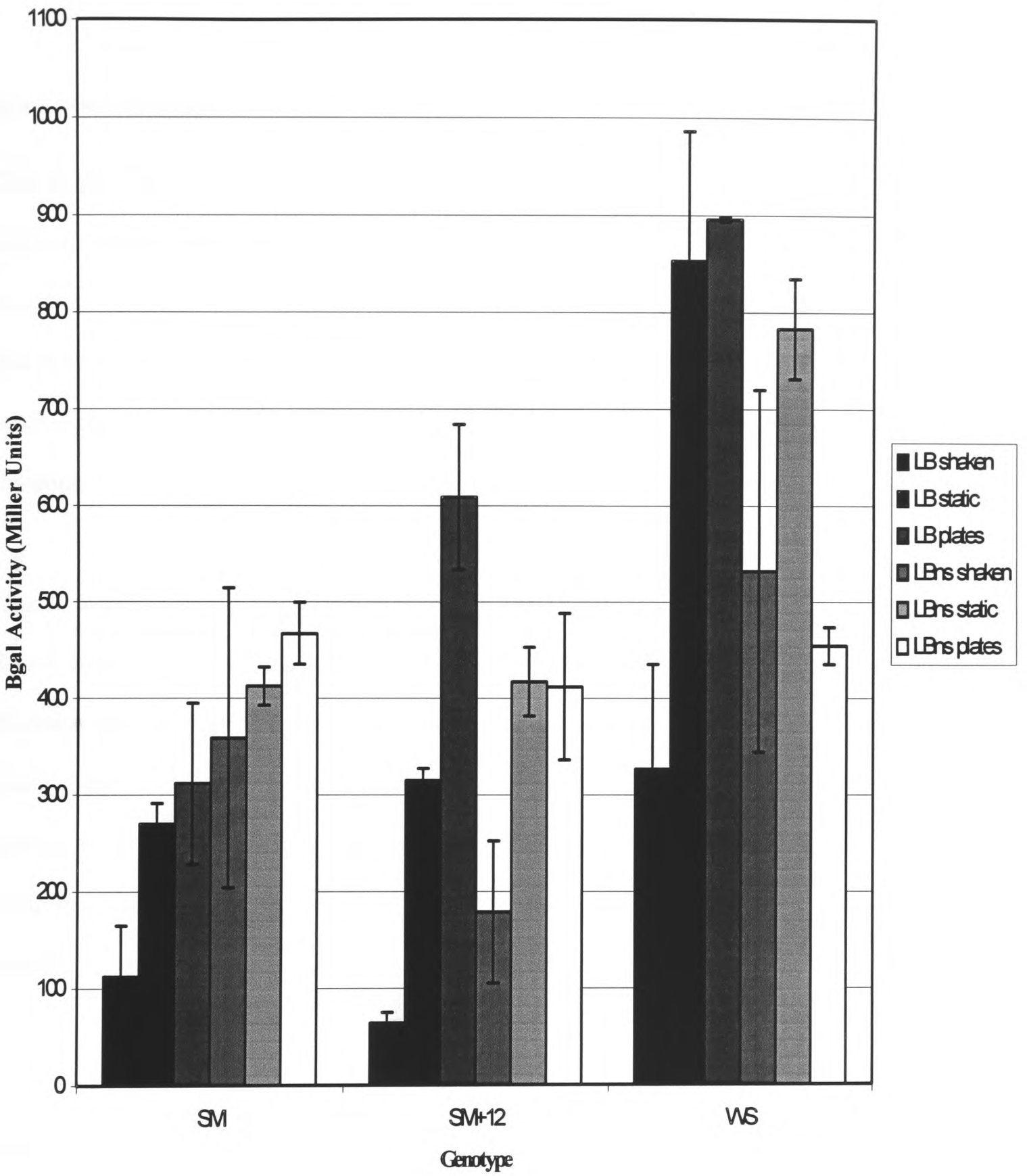


Figure 3.6. *Wss* Transcription under Different Environmental Conditions. The β-galactosidase activity of AS14 and AS15 containing pVSP61 (shown as SM and WS respectively to represent their phenotypes), and AS14 containing *wspR12* expressed from pVSP61 (the salt-dependent genotype) are shown under six different environmental conditions: spatially structured with salt (LB static); spatially unstructured with salt (LB shaken); agar plates with salt (LB plates); spatially structured without salt (LBns static); spatially unstructured without salt (LBns shaken); and agar plates without salt (LBns plates). The error bars are the standard errors from the three replicates performed in each case.

3.3 Discussion

3.3.1 The domains of WspR

The PILEUP analysis of WspR clearly shows the two separate domains of the protein, a result also supported by Bantinaki (2001). The role of domain shuffling in the evolution of response regulators has been discussed in Chapter 1 and it seems that *wspR* is an example. The fact that the domains have differing evolutionary origins makes the boundary between them clear, but it also means that two different sets of homologies need to be studied to gain any insights into function.

The N terminus shows homology to the input domains of many response regulators such as CheY from *E. coli*. As discussed in Chapter 1, these function due to phosphorylation by histidine kinases, such as WspE in *P. fluorescens*. WspR shares a conserved aspartate residue (amino acid no. 67 in this case) with these proteins and, in functional studies, this has been shown to be the phosphorylation site (Sanders *et al.*, 1989). This residue is an obvious candidate for site-directed mutagenesis in further studies. The fold analysis confirms the identification of this domain as the input part of a response regulator.

The two domains are separated by a linker region which has no obvious homologues. This is unsurprising due to its short length and the fact that there are probably few selective constraints acting on it. This does not, however, mean that it does not have a defined secondary structure and it is likely to be important in WspR function, especially in the light of allele 14 discussed below. Indeed the secondary structure analysis of *wspR19* and *wspR14* show that some helix formation is likely in the linker.

The C terminus has greatest homology to the *pleD* gene of *Caulobacter crescentus*; both these genes belong to a class characterised by the GGDEF motifs in their protein sequences that have been discussed in Chapter 1, especially with regard to their potential interaction with cyclic di-GMP. It is noteworthy that *wspR5* is mutated in this motif. The fold analysis of this domain is less likely to retrieve significant matches than that of the N-terminal domain because the crystal structures available do not include any very close homologues. However the highest hits that do occur suggest a nucleotide cyclase activity which is consistent with an interaction with cyclic di-GMP and also with the analysis of Pei and Grishin (2001).

3.3.2 The structure-function relationship of the *wspR* alleles

The combined results of the PILEUP analyses, the sequencing of the alleles, and the niche preference tests, show a striking correlation between the position of a mutation and its phenotypic effect. Those alleles with mutations in the N terminus or linker region (*wspR14* and *wspR19*) cause a transition from SM to WS when overexpressed, whereas the C-terminal mutants have precisely the opposite effect. The wild-type allele has a similar effect to the N-terminal mutants except for the salt-dependence which will be discussed below. It is also notable that all the alleles have mutations at sites which are highly conserved (for *wspR14* this is impossible to tell). This indicates that these residues are important for function so it is no surprise that mutations in them produce non-neutral phenotypes.

3.3.2.1 N-terminal mutants

The effect of overexpressing *wspR14* and *wspR19* is interesting from both mechanistic and evolutionary points of view. The transition from ancestral SM to LSWS is the evolutionary change of interest in this study, so any mutant allele causing this is potentially insightful. The discovery of these two alleles justify the study of *wspR* as a candidate gene because changes in it alone (albeit in expression levels as well as coding sequence) are sufficient to produce a WS phenotype, even if this is not what has actually occurred in the evolution of the LSWS

being studied. It is also of note that these alleles effect niche colonisation and also cellulose expression, confirming that they are using a similar route to the production of a WS phenotype; it is conceivable that there are non-cellulose routes to the surface-colonising phenotype.

The comparison between those alleles whose overexpression converts SM into WS under all the conditions studied, and the wild-type allele, whose overexpression produces a conditional WS phenotype, leads to their classification as overactive, dominant, or locked-on, alleles. They are potentially useful in studying the structural basis of *wspR* function as well as being useful tools for creating WS morphs and for *in vitro* studies of *wspR* activity due to their lack of requirement for other factors. They would, for example, be useful proteins to purify and use in cyclic di-GMP assays as they would function without the other components of the *wsp* operon.

Possible mechanisms for the interaction of N and C-terminal domains of response regulators have been discussed in Chapter 1. One of the most common is for the N terminus to repress the naturally-active C terminus until the N terminus is phosphorylated, when the repression is lifted. If this is the case in WspR it would seem likely that the mutations in *wspR14* and *wspR19* simply disrupt N-terminal function. This seems probable as the mutations would be loss of function at a local level and only gain of function at the level of the entire protein. Loss of function mutations are more likely to occur in large numbers. This also makes sense of the fact that *wspR14* is mutant in the linker region which is presumably involved in conveying the repression from one domain to the other.

The N-terminal mutants, when overexpressed, mimic the evolutionary transition from SM to WS in the way that some of the candidate gene studies mentioned in Chapter 1 did. However, as already mentioned, and as will be discussed further in Chapter 6, the mutation causing the transition is unlikely to be in *wspR* itself. This does not mean however that these mutants are

without evolutionary interest. They may well illustrate the type of mutation that causes the transition – it may be a mutation in a signalling protein that causes signal independence, possibly further upstream than WspR.

3.3.2.2 C-terminal mutants

The effect of the dominant alleles mentioned above, together with the work that implies that the SM to WS transition is a gain of function (Kahn, 1998, see also Chapter 1), makes the C-terminal mutants harder to classify. The effect they have on LSWS, turning it into a SM morph in all assays, is a dominant effect as it overrides the assumed activity of the chromosomal-encoded copy of wild-type WspR. It is, however, unlikely that it is a classical gain of function as this would mean that WspR can gain opposite functions by single base pair changes. While this is conceivable a more likely explanation, given the other evidence, is a model based on what is known about two-component signal transduction pathways and the effect of overexpression of certain non-functional components. Certain authors (Nakashima *et al.*, 1991; Morrison and Parkinson, 1994; Kim *et al.*, 2001) have found that the overexpression of liberated domains or non-functional copies of response-regulators leads to the quenching of the histidine kinase activity. This is illustrated in Figure 3.7. and is due to the stoichiometry of the functional and non-functional copies; the latter are essentially soaking up the input signal hence repressing the activity of the former. This will be integrated into a model of WspR activity given below.

3.3.3 WspR as a response regulator in a two-component system

As discussed at greater length in Chapter 1, two-component signal transduction is ubiquitous in prokaryotes. Typically the two components are a sensor protein, that detects an environmental signal, directly or indirectly, and a response regulator that transduces this signal, usually causing a change in gene expression or enzymatic activity. The sensor protein is a histidine kinase which, in the presence of the signal, autophosphorylates and then

phosphorylates the input domain of the response regulator. The response regulator catalyses this process, transferring the phosphoryl group from the histidine of the sensor to the conserved aspartate of its own N terminus. WspR has the conserved aspartate in its N terminus and is therefore thought to be a two-component response regulator. In such a model the N terminus would be the input domain, phosphorylated by an upstream sensor, probably WspE, while the C terminus would be the output domain, functioning when the N terminus is phosphorylated and affecting cellulose synthesis, possibly involving the GGDEF domain and cyclic di-GMP. This theory will be integrated into the discussion of the conditional phenotype of SM(pVSP61-*wspR12*) to produce a testable model of WspR activity.

3.3.4 Overexpression of wild-type *wspR*

The phenotypes produced by overexpressing the wild-type allele are similar to those produced by the N-terminal mutants: no effect on the LSWS and causing transition from SM to WS. The crucial difference is that this latter effect occurs only in the presence of salt. It cannot be that wild-type WspR requires salt directly for its function, as the chromosomal copy functions in LSWS when there is no salt. It is, however, possible that salt is a factor sensed by the signalling pathway of which WspR is a component, and that this salt signal is no longer required for *wspR* activity in LSWS due to an upstream mutation. If WspR is a rate-limiting step in the pathway then any increase in its amount could lead to increased activity of the pathway. The amount of *wspR* transcribed by the expression plasmid is probably at least 20-fold more than that produced by the chromosome so, as long as there is little translational regulation, and depending on the number of rate-controlling steps, there is potential for this increase in WspR copy number leading to cellulose production. An attempt to increase cellulose production in the ancestral smooth by increasing salt levels did not succeed. This may be due to the salt killing the cells at the concentrations that would be required, or WspR being at maximum activity on the initial amount of salt and therefore unable to increase output without an increase in copy number.

3.3.5 A model of WspR activity

These considerations of the results from this chapter allow a model of WspR activity to be proposed. The first part of this model is illustrated in Figure 3.7. In the LSWS genotype, which possesses a wild-type copy of *wspR*, it is thought that an upstream mutation has increased dramatically the signal being received by WspR and made its activity constitutive. As discussed above this mutation may be similar in type to those found in *wspR14* and *wspR19*, but in a different gene. Therefore cellulose is overproduced because WspR is always active, whilst this does not occur in the ancestral SM because WspR lacks a sufficient input signal. When the C-terminal mutants are overexpressed in the LSWS genotype they quench the constitutive signal as discussed above, and therefore the endogenous copy remains inactive and the cells revert to the ancestral state. If this is the case it is likely that liberated N termini will be able to have the same effect when overexpressed as they also will be able to soak up the signal. This would show that the precise sequence of the C-terminal mutant is not important, just the fact that it has lost its output ability while retaining its input one.

If the signal that activates WspR is indeed phosphorylation at the conserved aspartate residue number 67, then the removal of this residue should result in an allele that cannot be activated and that, if introduced into a LSWS cell in place of the wild-type allele, would not allow the production of cellulose. Contrarily, the N-terminal mutants are thought to be constitutively active because of the failure to repress the C-terminal output activity. These alleles would therefore function in any cell, SM or WS, turning on cellulose production. If this is the mechanism of N-terminal mutant activity then it may be that a liberated C terminus, not subject to N-terminal repression, could mimic one of these mutants. Such an effect is called jamming and has been observed for the C terminus of chemoreceptors (Ames and Parkinson, 1994).

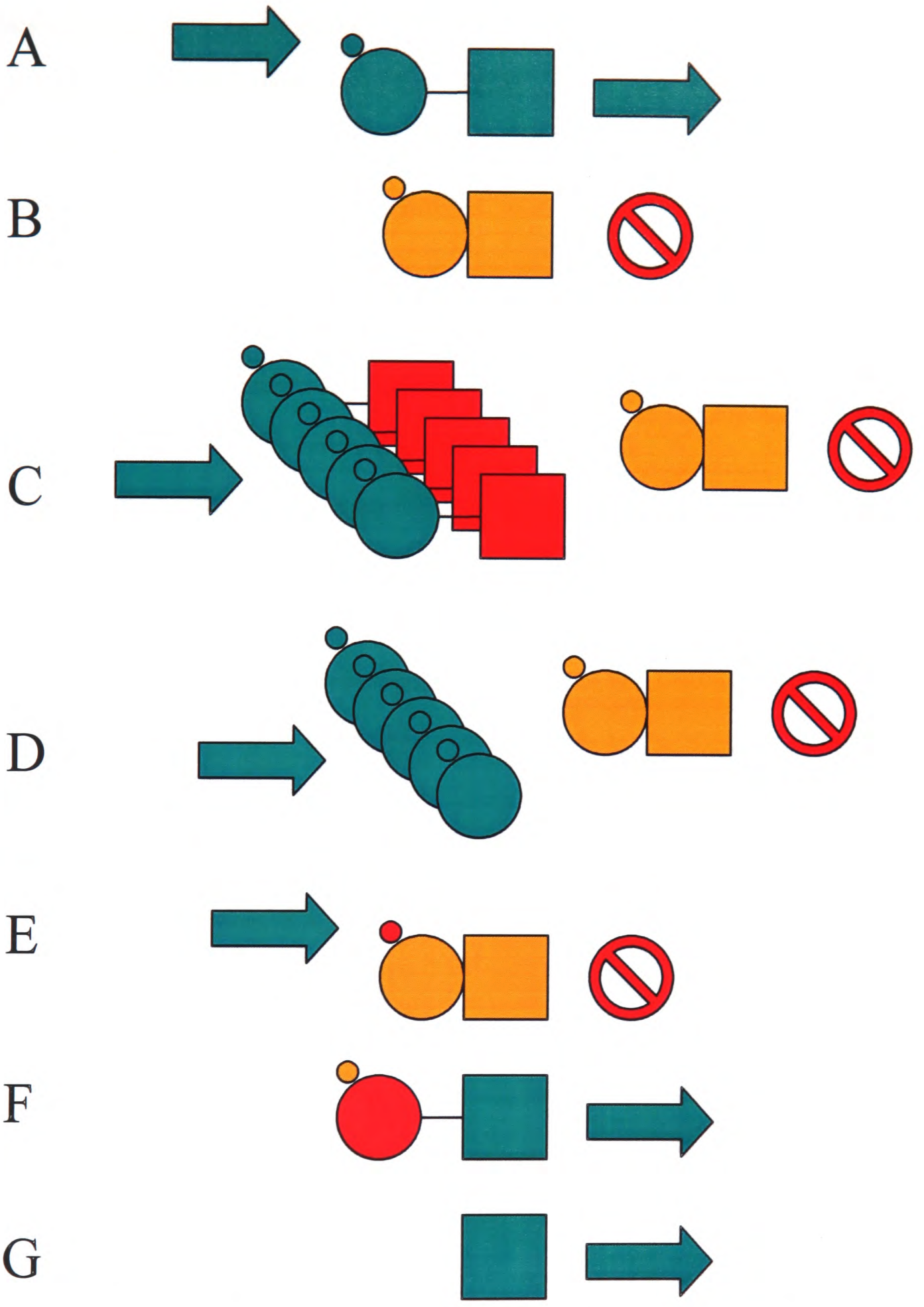


Figure 3.7. A Model of WspR Activity (Part 1). *WspR* is represented by a circular N terminus (with aspartate-67 shown separately as a small circle) and a square C terminus, along with input and ouput arrows. Throughout the diagram red domains are non-functional, orange are unactivated, and green are active. A number of scenarios are shown. Sections A and B represent the situations in the LSWS and SM, respectively. In the former, WspR is receiving a constitutive signal through phosphorylation. This causes the N terminus to release it's repression of the C terminus (represented by the termini moving apart), hence leading to activity. In the latter the lack of phosphorylation prevents this from happening. Section C shows the quenching of the signal by C-terminal mutants. The repression by the N terminus is lifted due to phosphorylation, but the C termini are inactive. Because of their saturating numbers the wild-type *wspR* does not receive the signal. The same is proposed to be the case with liberated N termini as shown in section D. Section E shows the likely effect of a phosphorylation site mutant: though the signal is present, it cannot be transduced. Section F shows a proposal for the activity of the N terminal mutants: though there is no signal the non-functional N terminus fails to repress C terminal activity as if it had been phosphorylated. The same is proposed to be the case for a liberated C terminus as illustrated in section G.

The second part of the model focuses on the effect of copy number on the wild-type allele, as seen with the conditional phenotype of SM(pVSP61-*wspR12*), and is illustrated in Figure 3.8. This depends on assigning magnitudes to the activities of WspR and the histidine kinase and making assumptions about the relative duration of their interaction compared with the time for which they remain active; the numbers given in the figure are examples. In the LSWS genotype it is assumed that the signal is sufficient for both WspR and the histidine kinase to be at maximum activity, hence cellulose is produced. In the ancestor, in the absence of salt, there is no activity of either. As salt is titrated into the medium the activity of the two components rise proportionally and would eventually produce a WS phenotype. However, the amount of output required for the WS phenotype is only produced by quantities of salt that are above the level that kills the cells. Hence, the WS phenotype cannot be induced in the ancestor in this way. This can be changed, however, when the relative amounts of histidine kinase and WspR are altered. If the latter is increased by several fold, it may be that each copy can reach the same level of activity as the histidine kinase, assuming that this can phosphorylate many copies of WspR which all retain their activity. The combined effect of all these partially active copies of WspR would be similar to that of the full activity of the wild-type number of copies found in the LSWS.

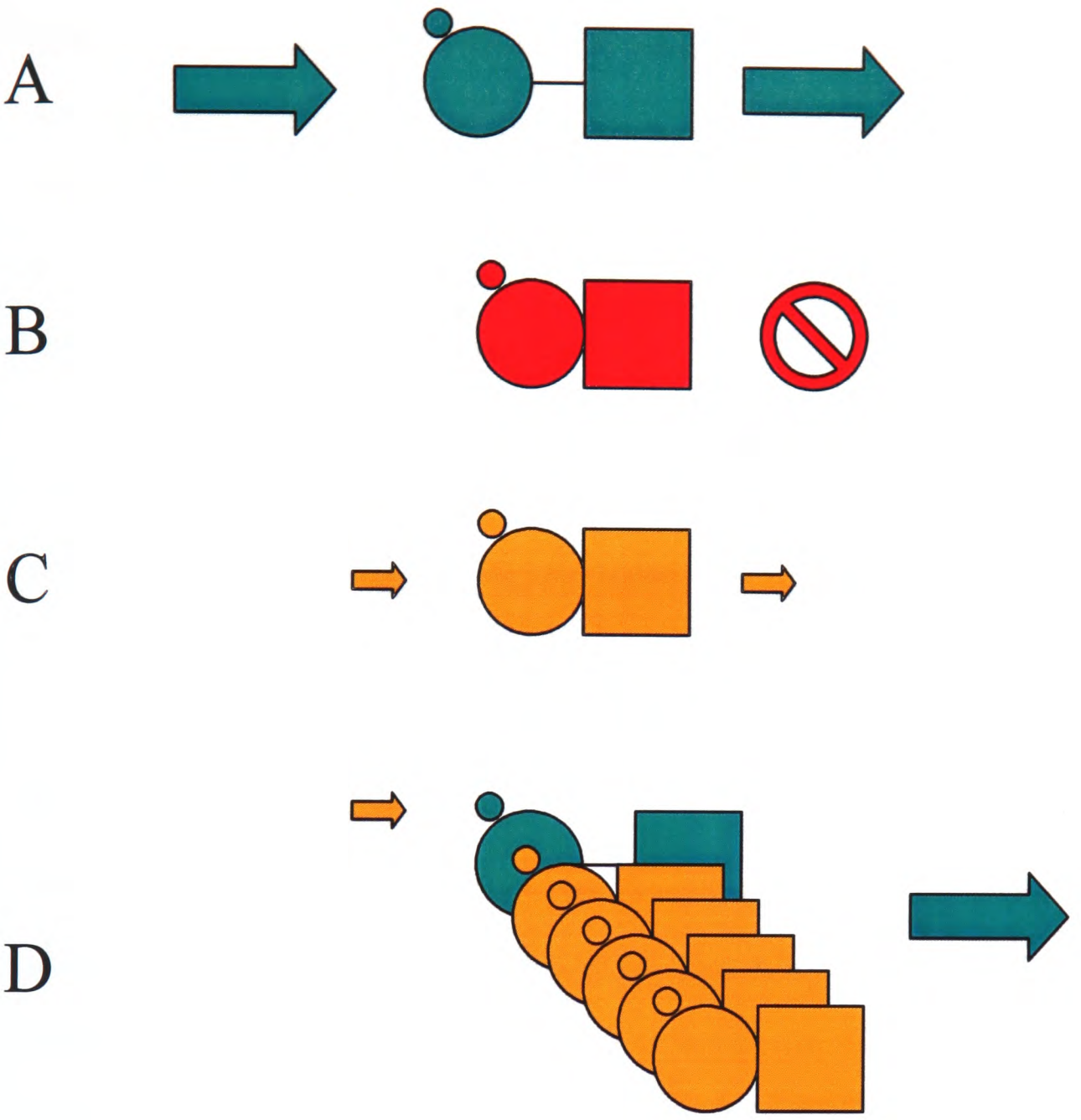


Figure 3.8. A Model of WspR Activity (Part II). WspR is represented as in Figure 3.7 in a number of scenarios. Throughout the diagram copies of WspR shaded green are fully active, those shaded orange are active 20% of the time and those shaded red are inactive. Section A represents the functioning of WspR in LSWS with a large input signal and a resultant large output signal. An arbitrary assumption is that each histidine kinase is active 100% of the time so that WspR is as well. Section B shows that there is no input signal into SM when there is no salt. When salt is present as in Section C the signal is increased and perhaps 20% of histidine kinases are active. The resultant 20% activity of WspR is not sufficient to turn on the WS phenotype. If the salt signal could be increased to 100% then it is imagined that the WS phenotype would result. However, the side effects of the salt are lethal. In Section D 20% of histidine kinases are active and this again results in each WspR being active 20% of the time. In this case, though, the 5-fold increase in WspR copy number causes a greater output leading to the WS phenotype. This model assumes that WspR is a rate-limiting step in the signalling pathway. It is also assumed that the increase in input signal in LSWS is caused by an upstream mutation that does not have the same lethal effects as an excess of salt.

This model assumes that the *wsp* operon is expressed at equal levels in both ancestor and LSWS. This has been shown at the level of transcription by Bantinaki (2001). It is possible that differences in the amount of translation of the proteins are in part responsible for the transition from ancestor to LSWS, which would add an extra component to the model. However, Bantinaki (2001) found no difference in the sequence of the operon in the two genotypes that would allow such an effect.

The predictions of this model, such as the activity of liberated domains and the effect of mutating the phosphorylation site are testable and will form the basis of the next chapter.

3.3.6 Transcription of the *wss* operon

The interpretation of the transcription data is not clear cut. First for consideration are the results illustrated in Figure 3.5 in which all the samples are measured in the same, spatially unstructured, shaken environment. This is not necessarily the ideal environment for the assays as the biofilm is not formed here by any genotype. It has been observed however that some degree of aggregation does occur when WS genotypes are cultured in shaken environments resulting in slower growth. The first observation is that there is a observable, but not substantial, difference in *wss* transcription between the ancestor and the LSWS. Also, four of the six genotypes with the WS phenotype [SM(pVSP61-*wspR14*), SM(pVSP61-*wspR19*), LSWS(pVSP61-*wspR12*) and LSWS(pVSP61-*wspR14*)] have higher transcript levels than the ancestor. This does not, however, demonstrate that WspR is affecting transcription, for two reasons. Firstly, the differences are very small and would require a non-linear relationship between transcript levels and cellulose production to account for the WS phenotype. Secondly, there is some possibility of positive feedback whereby the *wss* operon gene products increase their own transcription when more active. This is, however, a limited possibility as the *lacZ* fusion disrupts the genes from *wssE* onwards, so any effect seen here

would be due to the first four genes of the operon. There are two WS genotypes [SM(pVSP61-*wspR12*)and LSWS(pVSP61-*wspR19*)] that do not increase the transcription with respect to the ancestor which also put doubt on the possibility of a transcriptional effect. (These assays were performed with salt in the medium so SM(pVSP61-*wspR12*)is a WS genotype). The dominant negative alleles do, however, reduce all the *wss* transcription levels to that of the ancestor.

The most striking feature of the graph, however, is the threefold increase in transcription compared with all other strains seen in SM(pVSP61-*wspR19*). This is particularly odd given that this same allele reduces *wss* transcription in LSWS. It is hard to tell what might be happening here but the colony morphology and niche preference assays have already hinted that this is the allele with the severest effect. It may be that it increases transcription of *wss* by some unknown mechanism, but is so detrimental to the LSWS that it reduces levels of transcription in general. This detrimental effect of this allele will be considered further in Chapter 5.

The data in Figure 3.6 shows the effects of both salt and environment on the SM ancestor, LSWS, and the conditional genotype SM(pVSP61-*wspR12*). These effects are complex and both genotype and environment cause significant variation in *wss* transcription. Oddly, transcription tends to be higher in the absence of salt, even though this is where the conditional genotype is SM. It may be that salt causes a decrease in overall transcription, rather than in *wss* transcription in particular. There is an expected increase in transcription in static environments over shaken ones. The fact that the numerical values here are different from those in Figure 3.5 shows that there is variation between experiments, though it is hoped not in the relative values. Together, the two graphs show that there is more variation in one genotype between environments than there is between genotypes in one environment, though both effects are significant.

3.3.7 WspR output

The changes in cellulose production, and hence niche colonisation, caused by *wspR* alleles do not correlate well with differences in *wss* operon transcription, and, crucially, some genotypes have similar levels of transcription but opposite phenotypes. Other data has shown that an induced increase in *wss* transcription in the ancestor can cause an increase in cellulose production, but not a change in niche preference and colony morphology (J. Bohannon and P.B. Rainey, unpublished data). The combination of these two lines of inquiry suggest that WspR is acting post-translationally.

This is entirely consistent with the homology studies of *wspR* mentioned in Chapter 1. If the gene encodes a response regulator that acts through the nucleotide cyclic di-GMP then changes in transcription would not be expected. It is highly likely that cyclic di-GMP acts as it does in other systems, by interacting with the cellulose synthase complex to increase activity. In this case the overactive WspR mutants, and indeed the wrinkly spreader, would be as they are because WspR is producing or sequestering more nucleotide for the complex.

In the next chapter the theories derived from the results of this chapter will be tested more rigorously.

Chapter 4

Molecular Genetic Analysis of *wspR*

A model of WspR activity was proposed in Chapter 3 on the basis of the study of random point-mutants. In this chapter the predictions of this model are tested with the use of directed mutagenesis. This includes the use of point mutations to test the importance of particular residues, truncated proteins to test the function of different domains, and hybrid proteins to assess the degree of overlap in function with other systems

Aims:

1. To determine the role of phosphorylation in WspR function
2. To assess the basis of the signal-independent phenotype of certain alleles
3. To discover if multiple non-functional copies of WspR are capable of quenching a signal
4. To determine the epistatic interactions of *wspR* with other genes
5. To find out if WspR and PleD can exchange functions

The work performed in this chapter in *Caulobacter crescentus* was kindly carried out by Phillip Aldridge at The University of Basel, Switzerland.

4.1 Introduction

In Chapter 3 *wspR* was analysed by way of a set of alleles obtained from a process that amounted to random mutagenesis. The insights gained from this approach allow specific questions to be asked about the structure and function of the gene which can be answered by directed mutagenesis. The questions to be answered include the roles of the putative phosphorylation site in the N terminus and the conserved GGEEF motif in the C terminus, the suggested quenching effect of the C-terminal mutants discussed in the previous chapter, the potentially epistatic interactions of *wspR* mutants with mutations in other genes, and the functional homology between WspR and PleD.

The PILEUP analysis of the WspR N terminus in Chapter 3 (Figure 3.1.) shows a universally conserved aspartate residue at position 67 which is predicted to be the site where the response regulator is phosphorylated by the histidine kinase. In the model proposed in the previous chapter this signal causes a conformational change that lifts the repression of the C terminus by the N terminus. A genetic approach to establishing whether phosphorylation is involved in WspR is to mutate the aspartate residue to an asparagine thus removing the possibility of phosphorylation while retaining the structure as closely as possible. Aspartate and asparagine are virtually identical in size and structure, but radically different chemically, the carboxyl group of the former being replaced by an amide group in the latter. Therefore the only difference caused to WspR by this mutation would be the loss of its ability to be phosphorylated. If the protein loses its function it is likely that phosphorylation is necessary for activation, while if it becomes constitutively active then phosphorylation is probably required for deactivation.

The PILEUP analysis of the *wspR* C terminus (Figure 3.2.) illustrates well that the GGEEF motif is highly conserved (though it is GGDEF in many other cases), and work in *A.xylinum* (Tal *et al.*, 1998) implicates it in the interaction with cyclic di-GMP, a regulator of cellulose

synthesis. One of the WspR mutant alleles studied in the previous chapter (*wspR5*) contained a serine residue in place of the phenylalanine of the GGEEF motif, hence further site-directed mutagenesis of this motif is unnecessary because it has already been shown to be necessary for function.

As discussed in the previous chapter, it is thought that the dominant negative effect of the C-terminal *wspR* mutants on the LSWS strain is due to quenching of the signal by multiple non-functional copies of the protein. It was also postulated in Figure 3.6. that a similar effect should be possible by overexpressing a liberated N terminus, provided that the protein is of a modular nature, with the N terminus not relying on the presence of the C terminus for uptake of its signal. A related question, also postulated in the model of WspR activity, is whether the C terminus is constitutively active and is suppressed by a dephosphorylated N terminus. If this is the case, and the C terminus does not rely on the presence of the N terminus for its function, then a liberated C terminus should have the same effect as a constitutive N-terminal mutant, referred to as jamming. An attempt will be made in this chapter to create both these effects.

The introduction of *wspR* alleles into genotypes carrying mutations engineered into other genes allows an analysis of sufficiency and necessity for the different components in the pathway leading to the adaptive WS phenotype. A classical genetic study based on epistasis has been carried out in several cases.

Finally, the close homology between WspR and PleD in their C termini (illustrated in Figure 4.1.) suggests that there may be functional equivalence. This can be tested by attempts at cross-reactivity using both native genes and hybrids. A similar study of functional interchangeability between GGDEF proteins has been performed by Ausmees *et al.* (2001). However, these authors used only wild-type genes and not hybrids; by using hybrids, input and output signals into the protein can be manipulated separately.

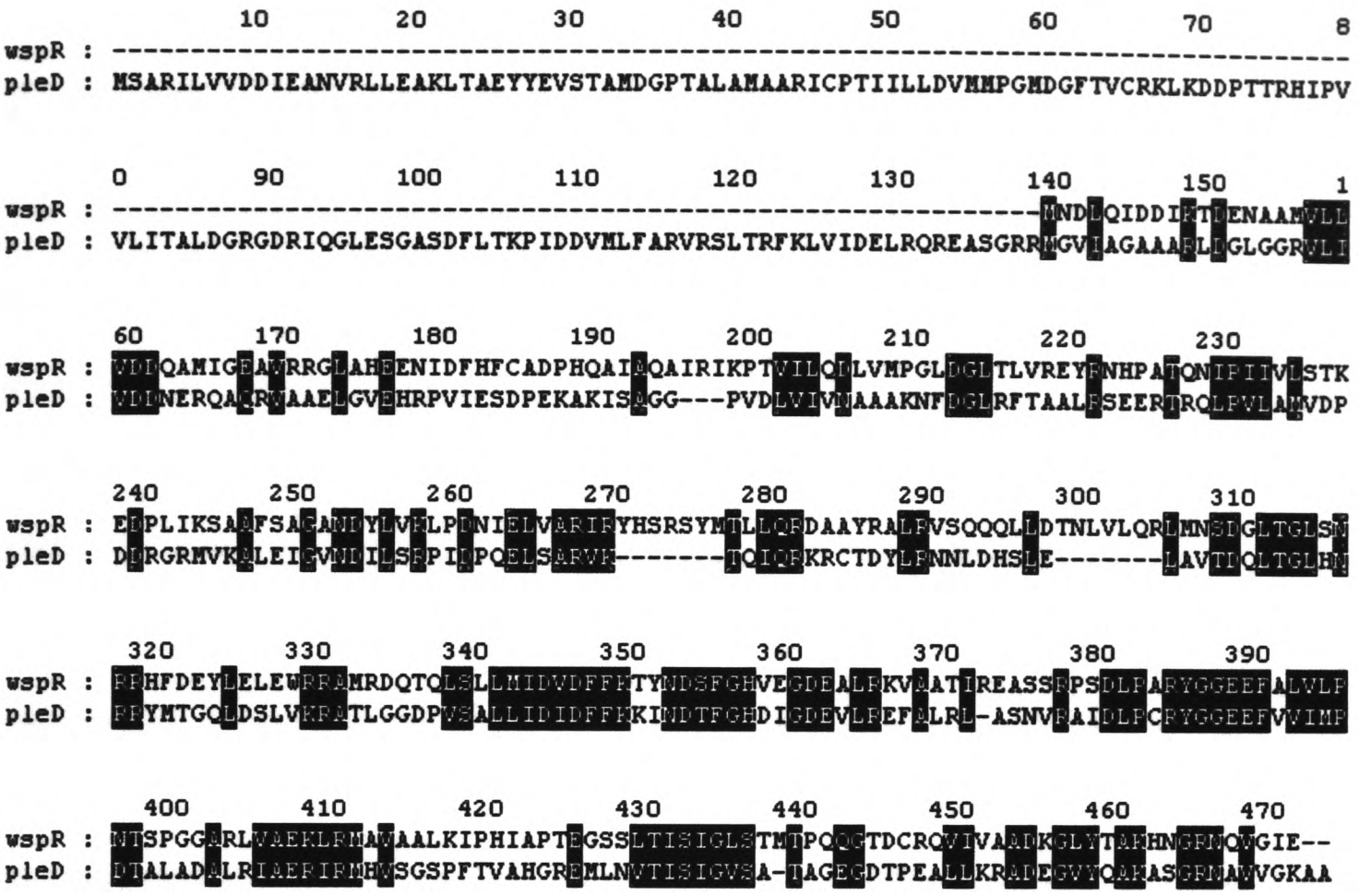


Figure 4.1. Pairwise Alignment of WspR and PleD C-terminal Protein Sequences. Sequences were paired using BLASTp 2 and the PAM250 matrix. Both similar and identical residues are shaded. A second alignment, between WspR N terminus and the first N-terminal domain of PleD is not shown.

4.2 Results

4.2.1 WspR function depends on phosphorylation

The conserved aspartate residue at position 67 of WspR is a phosphorylation site in those homologues for which function has been assigned (Sanders *et al.*, 1989). To establish whether phosphorylation was also necessary in WspR this was mutated to an asparagine residue, which maintains the structure of the protein as closely as possible while removing the chemical possibility of phosphorylation. Primers *wspR*-DN-for and *wspR*-DN-rev were used, along with *wspR*-for and *wspR*-rev, to produce a mutant PCR product by SOEing. In this approach two overlapping parts of *wspR* are produced containing the mutation in the overlap due to its presence in the primers. These two fragments are then put together by strand overlap extension (for full details see Chapter 2). The sequence of the primers replaces the GAT of an aspartate codon with the AAT of an asparagine one. The mutant *wspR* PCR product (designated *wspR12d*) was cloned into pVSP61 using the same combination of primers and restriction enzymes as was used for the original *wspR* alleles (Kahn, 1998; also Chapter 1 of this study). For positions of all the mutations created in this chapter see Table 4.1.

Allele	Changes to *wspR* nucleotide sequence	Changes to WspR protein sequence	First amino acid of WspR in mutant	Last amino acid of WspR in mutant	First amino acid of PleD in mutant	Last amino acid of PleD in mutant
wspR12d	199, g to a	67, D to N	1	333	N/a	N/a
wspR19d	199, g to a 385, c to t	67, D to N 129, R to C	1	333	N/a	N/a
wspR4	385, c to t 677, a to c	129, R to C 226, L to S	1	333	N/a	N/a
wspR-NtermA	N/a	N/a	1	136	N/a	N/a
wspR-NtermB	N/a	N/a	1	141	N/a	N/a
wspR-Cterm	N/a	N/a	158*	333	N/a	N/a
WspR19-pleD1	385, c to t	129, R to C	1	154	276	454
WspR19-pleD2	385, c to t	129, R to C	1	170	291	454
pleD-wspR	N/a	N/a	171	333	1	292
pleD-wspR9	886, g to a	N/a	171	333	1	292

Table 4.1. *wspR* and pleD mutants. The point mutations, liberated domains, and hybrid genes created in this chapter are all shown. The positions of mutations and the resultant change in protein sequence is given, as well as the positions in the protein sequence of WspR where the liberated domains begin and start, and in the sequences of both proteins where the hybrids are joined. *This liberated domain begins with a methionine followed by residue 158 of WspR.

The plasmid, pVSP61-*wspR12d*, was conjugated into SM and LSWS genotypes and the colony morphology and niche preference were scored. In neither case did the allele cause any change in the parental phenotype, indicating a loss of function. This is in contrast to the effects of both the wild-type which functions in the SM causing a phenotypic transition to WS, and the non-functional C-terminal mutants which have a dominant negative effect on LSWS. It is also in contrast to the other N-terminal mutants which have a dominant effect, despite also having a mutation in the N terminus, suggesting the unique importance of this particular aspartate residue. The phenotypic effect of this allele, and all the others described in this chapter, is shown in Table 4.2.

Allele	Phenotype when overexpressed in:		
	SM	**LSWS**	***C. crescentus ΔpleD***
wspR12d	SM	WS	N/a
wspR19d	WS*	WS	N/a
wspR4	SM	SM	N/a
wspR-NtermA	SM	WS	N/a
wspR-NtermB	SM	SM	N/a
wspR-Cterm	SM	WS	N/a
wspR-pleD1	SM	WS	No complementation
wspR-pleD2	SM	WS	No complementation
pleD-wspR	SM	WS	Complementation
pleD-wspR9	SM	WS	No complementation

Table 4.2. The phenotypes of *wspR* and *pleD* mutant alleles. The phenotypic consequences of overexpression in *P. fluorescens* and *C. crescentus* is shown for each allele described in Table 4.1. Phenotype refers to colony morphology for *P. fluorescens* genotypes SM and LSWS, and the ability to complement a *pleD* mutant for *C. crescentus*. * this allele, overexpressed in SM, shows a colony morphology intermediate between that of LSWS and that of SM.

4.2.2 WspR19 does not require phosphorylation

As discussed in the previous chapter, it is thought that *wspR19* and *wspR14* are signal-independent, locked-on, alleles. For the experiments in this chapter, *wspR19* was selected to represent the two as it has the severest phenotype and is therefore more likely to produce observable results. If they are indeed signal-independent, and if the signal in question is phosphorylation at residue 67, then carrying both the *wspR19* mutation and the aspartate-to-asparagine mutation should not result in loss of function, as the *wspR19* mutation should be epistatic to the loss of the phosphorylation site. Therefore such an allele (designated *wspR19d*) was cloned into pVSP61 using exactly the method used above to create *wspR12d*. Conjugation into LSWS had no effect on colony morphology, as expected. Conjugation into SM resulted in an intermediate phenotype on agar plates and biofilm formation in microcosms. This shows that the loss of the phosphorylation site from WspR19 does diminish its function. However, viewed the other way round, the addition of the *wspR19* mutation to a phosphorylation-deficient null-mutant results in a partial restoration of function, pointing to signal independence. The full implications of this will be discussed below.

4.2.3 A C-terminal mutation is epistatic to *wspR19*

If the explanations given for the phenotypes of the N and C-terminal mutants in the previous chapter are correct, the introduction of a C-terminal mutation into *wspR19* should abolish function and cause dominant negativity. This would be in contrast to the effect of introducing a phosphorylation mutation into *wspR19* as shown above. In order to do this with greatest ease a PCR cloning of *wspR19*, followed by a mass conjugation, were performed in exactly the same way as was used to isolate *wspR1* in Chapter 3. One of the dominant-negative clones isolated from this procedure was sequenced (as in Section 3.2.3) and the resultant allele called *wspR4*. This allele has the *wspR19* sequence combined with a T to C mutation at base 677 (causing a leucine to serine amino acid change at residue 226) in the C terminus. This demonstrates that an allele that combines a dominant mutation in the N terminus, with a dominant-negative mutation in the C terminus, when overexpressed, changes the colony morphology of LSWS to SM but has no effect on the morphology of the ancestral SM. Hence the C-terminal mutation in epistatic to the N-terminal one

4.2.4 A liberated N terminus can imitate a C-terminal mutant

The quenching effect thought to occur due to copies of WspR with non-functional C termini, may also be achievable with liberated N termini (as discussed in Chapter 3 and above). To determine this, two different truncated alleles were created consisting of residues 1-136 (*wspR*-NtermA) and 1-141 (*wspR*-NtermB) (See Figure 4.1). Two different versions were used to increase the chance of achieving a positive result if local structural considerations affected function. These were cloned into pVSP61 from PCR products created with primers *wspR*-for, and *wspR*-midrevC and *wspR*-midrevD respectively. Both the reverse primers introduced a stop codon after the last desired codon of *wspR*. The pVSP61 clones were then conjugated into SM and LSWS genotypes. Neither allele had an effect on SM, and *wspR*-NtermA had no effect on LSWS. However, *wspR*-NtermB caused the reversion of LSWS to a SM, non-biofilm forming genotype, an effect identical to that of the C-terminal mutants.

Hence the action of liberated N termini is dependent on their size, with the larger of the two truncated versions having a dominant-negative effect.

4.2.5 A liberated C terminus is non-functional

The previous result shows that the N terminus of WspR can fulfil at least some of the function of the native protein (receiving a phosphorylation signal) in a liberated form. It was therefore decided to test the effect of a liberated C terminus too see if a jamming phenotype was observed. This would only occur if the C terminus is naturally active, depending on the N terminus for repression when there is no input signal. Also there would need to be no structural requirement for the presence of an N terminus for C-terminal function. The C terminus was cloned into pVSP61 as above, using primers *wspR*-midforC and *wspR*-rev. This resulted in a truncated allele beginning with a methionine and then residue number 158 of WspR. This allele was overexpressed in both SM and LSWS genotypes, but no effect was observed. Given the stringent conditions for function it is not necessarily surprising that no effect was seen, though this will be discussed further below.

4.2.6 The effect of *wspR* alleles on *wsp* mutants

While WspR is sufficient to cause the transition from SM to WS in a wild-type cell, given the presence of salt, it is unlikely that it is sufficient when other components of the *wsp* operon are missing. To test this assumption both *wspR12* (the wild-type allele that has a salt-dependent effect on the ancestral SM) and *wspR19* (the locked-on N-terminal mutant that changes SM into WS regardless of salt concentration) were overexpressed in WS4, a *wspR* mutant possessing the other components of the *wsp* operon, and in SMΔ*wsp* and LSWSΔ*wsp*, respectively SM and LSWS genotypes lacking genes *wspA* to *wspF* but containing *wspR* with an intact promoter (Bantinaki, 2001). In order to conjugate into WS4 the new, tetracycline-resistant, pVSP61-ΩTc vectors were constructed, as WS4 and pVSP61 are both kanamycin

resistant. These vectors were constructed by cloning the ΩTc cassette (Fellay *et al.*, 1989) as an *Eco*RI fragment into the *Eco*RI site at the end of *wspR*. As expected both alleles complemented the WS4 mutation in colony morphology and niche preference assays. The *wspR12* allele, however, does not complement the Δ*wsp* mutations, indicating its requirement for the other components of the *wsp* pathway. The *wspR19* allele, on the other hand, does cause the SM to WS transition in the Δ*wsp* genotypes, indicating its signal independent nature. In all cases the plasmid controls (pVSP61-ΩTc vectors containing no *wspR* insert) did not cause a visible change in colony morphology.

4.2.7 The effect of *wspR* alleles on *wss* mutants

4.2.7.1 WS13

WS13 is a *wssB* mutant, hence it is lacking a component of the cellulose synthase complex. As discussed in Chapter One, it was isolated from the initial mutant screen of LSWS and classified as a structural mutant that blocked the possibility of WS evolution (diversification occurred but no wrinkly spreaders were produced). It was subsequently mapped to the *wss* operon, encoding the cellulose synthase genes. The *wspR* alleles, *wspR12* and *wspR19* were conjugated into WS13 on the pVSP61-ΩTc vectors (as WS13, like WS4, is kanamycin resistant). In neither case was there complementation of the WS phenotype. This confirms that *wspR*, even in the form of *wspR19*, is not the end of the pathway producing the phenotype and that cellulose production is necessary.

4.2.7.2 JB01

JB01 is an ancestral SM genotype that has had a constitutive promoter, NPTII, inserted upstream of, and in frame with, the *wss* operon (Spiers *et al.*, 2002). That this alone does not result in a WS phenotype indicates that the regulation of cellulose synthesis is not just at the transcriptional level (see Chapter 3 for discussion). It does, however, result in an increase in cellulose production; it is unclear whether the amount of cellulose produced is insufficient for

the WS phenotype, or whether a qualitative difference is required such as modification of the polymer.

The *wspR* alleles *wspR9*, *wspR12* and *wspR19* were all conjugated into JB01 on their original pVSP61 vectors. The latter two resulted in a WS phenotype on agar plates whereas the former did not. The results for *wspR19* and *wspR9* are similar to those observed with the same alleles in the ancestral SM, namely no effect with *wspR9* and a SM to WS transition with *wspR19*. On the other hand the result with *wspR12* is different from that in the ancestor as in this case the SM to WS transition does not depend on the presence of salt. This may be due to the rate-limiting nature of WspR being altered by a different balance of *wss* and *wsp* operon levels.

4.2.8 The effect of *wspR* alleles on *Caulobacter crescentus*

As discussed in Chapter One, *C. crescentus* contains a *wspR* homologue, *pleD*, that encodes a response regulator required for stalk ejection during its complex life cycle. It is of interest to know if the structural homology between the two proteins extends to functional homology. The first step in answering this question was to see whether *wspR* could compliment a *pleD* mutant of *C. crescentus*. The pVSP61 plasmids containing *wspR12* and *wspR19* were conjugated into *C. crescentus* strain UJ284 (a *pleD* mutant) and phenotype was assayed. This was done by counting the number of cells that successfully eject their flagella under the microscope using the method of Aldridge and Jenal (1999) as described in Chapter 2. It was found that *wspR12* did not complement the mutation whereas *wspR19* did.

4.2.9 A hybrid of PleD and WspR

The fact that WspR19, which is predicted to be signal independent, functioned in *C. crescentus*, whereas WspR12, that requires a signal, did not, leads to the hypothesis that the nature of the input signal is where WspR and PleD differ, while the output domains are

interchangeable. WspR19 can function because it does not require a signal to activate its output domain, whereas WspR12 is not activated in *C. crescentus* and therefore does not function. If WspR12 could be provided with an input signal it would therefore be expected to function. As the signal received by the N-terminal domain of PleD is definitely present in *C. crescentus,* a fusion between this and the C-terminal domain of WspR might be active. Such a hybrid was constructed by SOEing and cloned into pPA100. This expression vector was found to complement the *pleD* mutant thus confirming the hypothesis. As expected the similarly constructed hybrid of PleD and WspR9 did not function. Neither of these alleles produced an effect in SM or LSWS genotypes of *P. fluorescens*.

4.2.10 The effect of PleD on *P. fluorescens*

If the output domain of WspR is able to mimic PleD in *C. crescentus* then it seems likely that PleD will similarly mimic WspR in *P. fluorescens*. As an initial test of this, *pleD*, cloned in the expression plasmid pPA41, was conjugated into the ancestral SM and LSWS. No observable colony morphology changes were seen in either case, though this is not surprising as the input signal for PleD is unlikely to be present in *P. fluorescens*. For this reason it was decided to construct the reverse hybrid of the one tested above. In this case the N terminus of WspR19 (the signal-independent form) was coupled with the C terminus of PleD by SOEing the genes into pVSP61. The external primers used were *wspR*-forA for the forward reaction on *wspR,* and *pleD*-revA for the reverse reaction on *pleD*. The former has the same sequence as *wspR*-for, used in the construction of the expression plasmids in Chapter 3, the only difference being that a *Hind*III site was used rather than a *Bam*HI site because *pleD* C terminus contains a *Bam*HI site. Two different pairs of internal primers were used to make two different hybrids, respectively *wspR*-midrevA and *pleD*-midforA, and *wspR*-midrevB and *pleD*-midforB. The former results in a hybrid connecting WspR amino acid number 154 to PleD number 276 (hybrid 1), while the latter connects WspR amino acid number 170 to PleD number 291 (hybrid 2). Hybrid 1 contains the linker region from PleD while hybrid 2 has the

sequence of WspR up to the beginning of the C terminus where the significant homology between the two proteins starts (see Figure 4.1.). The two hybrids were conjugated into SM and LSWS and no effect on colony morphology was observed in either case. Two alleles were used to increase the chances of a positive result. Hybrid 2 was considered the more likely to prove successful as the C terminus of PleD replaced that of WspR at the least obtrusive point. Likewise, no effect of these alleles was seen in the *C. crescentus pleD*-mutant strain UJ284.

4.2.11 Signal-independent *pleD* alleles

The failure of the WspR19-PleD hybrids to cause the SM to WS transition led to the possibility that the WspR N terminus is not compatible with the PleD C terminus. If this is the case, and wild-type PleD does not function in *P. fluorescens* because of the lack of input signal, the only possible way to get PleD to mimic WspR is to use a signal-independent mutant allele. Such an allele would be the *pleD* equivalent of *wspR19*. Several such mutants were obtained by random mutagenesis and cloned into vectors pPA11428, pPA11433, pPA11435, pPA11441, pPA11443 and pPA11447. They were isolated on the basis of their effect in *C. crescentus* and their sequence is unknown. These were conjugated into SM and LSWS, but no changes in colony morphology were observed.

4.3 Discussion

4.3.1 Phosphorylation

The substitution of the aspartate residue at position 67 of WspR with asparagine (allele *wspR12d*) and its subsequent lack of function show the necessity of this residue. The fact that this loss of function was observed even though the mutation was to asparagine, structurally virtually identical to aspartate, implies that the requirement is a chemical one. The conservation of this residue in all the N-terminal homologues, and the fact that it has been shown to be phosphorylated in biochemical studies (Sanders *et al.*, 1989) is very strong evidence for its phosphorylation being required for WspR function. The only evidence that would be even stronger would be to have actual biochemical data, requiring purified protein or partially purified cell extract.

The unique significance of the aspartate residue is also indicated by the mutant allele (*wspR12d*) causing a phenotype that is different from the other ones studied so far. While it is non-functional, it does not have the dominant-negative phenotype of the C-terminal mutants. On the other hand, although it is an N-terminal mutation, it does not have the dominant phenotype of WspR14 and WspR19. This lack of any function is consistent with this mutant having lost the ability to receive its input signal so that, even when overexpressed, it is not able to quench like a C-terminal mutant. It does, however, pose the problem of interpreting a completely negative result. The lack of an antibody to WspR means that Western blots to check that the protein is being expressed are not possible. It has, however, been cloned in an identical way, with the same primers and restriction sites, to all the other *wspR* alleles which have produced visible phenotypes, so there is no reason to think that it is not present. In addition to this it should be noted that overexpression of *wspR19d* produces a detectable phenotype and was cloned in tandem with *wspR12d*, differing at only one base. This

precludes the possibility that the aspartate-apsaragine transition affects the transcription or translation of the gene product.

4.3.2 Signal independence

The aspartate to asparagine mutation causes a reduction in the severity of the WspR19 phenotype. It does not, however, produce a conditional phenotype like that caused by the overexpression of *wspR12*. The effect can be viewed in the opposite manner: the non-functional phosphorylation mutant (WspR12d) undergoes a gain of function on the introduction of the *wspR19* mutation. This functional rescue is different from that gained by reintroducing the aspartate in that it is signal-independent. There are several possible reasons why WspR19d has a reduced activity: It may be that the aspartate does play a structural role which asparagine cannot fulfil, though this is highly unlikely; alternatively WspR19 may promote its own phosphorylation when there is no signal to further increase its activity, such an effect being impossible in WspR19d. There is evidence (Bourret *et al.*, 1990; Appleby and Bourret, 1999) that CheY can be phosphorylated at alternative sites. It could be that WspR19 functions by promoting its own phosphorylation regardless of histidine kinase activity. In WspR19d this would only be possible at the alternative sites hence the reduction in activity. However, it would seem strange that WspR12d is not able to do this, although this could be a property of the histidine kinase. Either way these results together confirm the signal-independence of WspR19 (and possibly the explanation given for this in Chapter 3), and also the role of the aspartate residue as the signal recipient.

Crucially the combination of *wspR19* and a C-terminal mutation, unlike that of *wspR19* and the phosphorylation-site mutation, encodes a non-functional protein. This shows that the output domain is genetically epistatic to the input one as would be expected in the model of WspR as a response regulator.

4.3.3. Domain liberation

4.3.3.1 N terminus

The attempts to imitate the purported quenching effect of the C-terminal mutants by using a liberated N terminus gave mixed results. Of the two N termini that were tried only one had the desired effect. The crucial fact, however, is that one of them worked and this would still stand if many other deletions were tried and failed. Both versions covered the entire N terminus, as defined by conservation in Figure 3.1, the differences between them being in the amount of the linker region that remained. It is not hard to imagine that the structural integrity of the domain could depend on having certain linker residues, or possibly on not having certain ones. A comprehensive study involving a series of alleles that were each one codon bigger than the last might uncover a complex relationship between length and quenching ability because of precise structural considerations. In this case it was the longer of the two alleles that encoded a quenching domain but it could conceivably have been the other way round. Importantly, however, the fact that one of the two variants used produced the dominant-negative phenotype confirms the quenching interpretation.

There are two ways in which this quenching effect could be happening: The non-functional copies could be binding to the histidine kinase, preventing the wild-type from doing so, or the non-functional copies could be being phosphorylated, hence leaving insufficient kinase activity to phosphorylate the wild-type. One test for this would be to overexpress the same N-terminal domain that caused quenching, but with a phosphorylation site mutation. However, the fact that the phosphorylation mutant, WspR12d, does not have a dominant-negative phenotype when overexpressed from pVSP61, is sufficient to make the distinction. If binding to the kinase was being blocked then this would be expected to quench the signal when overexpressed. This is because there would be multiple non-functional copies that would bind the kinase just as the C-terminal mutants would. However, this was not found to be the case. WspR12d must not be able to quench because it cannot soak up the phosphorylation signal.

This favours the second interpretation that quenching is due to excessive uptake of the phosphorylation signal by the non-functional copies.

4.3.3.2. C terminus

Only one attempt was made to mimic an N-terminal mutant with a liberated C terminus (a jamming effect) so it is hard to draw many conclusions from the experiment. The requirements for a liberated N terminus to take up the phosphorylation signal are an accessible aspartate residue and a conformation that allows the kinase and response regulator to come into contact in a roughly similar way to that in the wild-type. It can be imagined that these requirements are fairly broad. The requirements for a liberated C terminus to produce the appropriate output might be considerably more specific. If an interaction with cyclic di-GMP, possibly an enzymatic one, is necessary, then several residues from different parts of the primary sequence would need to align correctly. It is conceivable that the absence of the other domain of the protein could affect such active sites considerably, especially if dimerisation is required for function. For this reason it is not considered surprising that the liberated domain was not functional, and no further attempt was carried out. It is impossible to tell whether there was no effect for the reason given above, or because the N terminus is not involved in suppressing a naturally active C terminus in the case of WspR.

4.3.4. Complementation of *wsp* and *wss* operons

The complementation studies of *wss* and *wsp* operons constitute a classical study of genetic epistasis, and confirm several conjectures about the two operons. The WS4 mutation (Kahn, 1998) showed that WspR was necessary for the WS phenotype. Bantinaki (2001) produced the LSWSΔ*wsp* deletion genotype that showed that at least some of the other genes in the *wsp* operon were also necessary. In this chapter it was shown that the equivalent SM genotype, SMΔ*wsp*, can be converted into a wrinkly spreader by the signal-independent WspR19, but not by the wild-type WspR12. This strongly implies that the genes of the *wsp* operon do not

just encode factors that are necessary for WspR function (in that case WspR19 would be expected not to work) but they encode the signalling pathway that actually feeds into WspR. This was expected on the basis of homology, but it was possible that WspR received a signal from another pathway.

This confirmation that the *wsp* operon is upstream of WspR in the pathway, and hence that WspR is dependent on *wsp*, is mirrored by the results with the *wssB* mutant WS13. As expected WspR19 does not complement this mutation, showing that it is downstream, though of course the mechanism by which WspR activates cellulose synthesis is still unknown. This result was also to be expected as the *wss* operon encodes the actual structural basis of the phenotype. This issue of factors upstream and downstream of WspR will be returned to in the proteomic approach used in Chapter 7.

The results seen when *wspR* alleles were overexpressed in JB01 are less easy to explain. It was expected that WspR12 would still produce a conditional effect in this genotype as it is still phenotypically SM. However the result is a WS phenotype. The model proposed in Figure 3.7 depends heavily on the relative amounts of different protein and signalling components. It is quite possible that this is also the case here and that the increase in *wss* operon components has led to the sensitivity of the system being radically altered.

All these genetic experiments prove useful points but they may well be pushing towards the limit at which the genetic approach will provide no more insights into the function of this system. This does not of course include genetic studies looking at complementation between species, as with *pleD*, or between different WS genotypes, as will be examined in Chapter 7. It may, however, be that biochemical studies are now needed to learn more about the role of WspR in LSWS.

4.3.5. Complementarity of WspR and PleD

The study by Ausmees *et al.* (2001) showed that some members of the GGDEF protein family could replace each other in the role of cellulose synthesis activation. It is probable that WspR would also have this activity, especially considering its very close homology to CelR2, one of the proteins used in the study. Those authors consider the homology between PleD and CelR2 and show that the latter does not appear to have a role in flagellar function. However, they do not make an attempt to see if there is functional complementarity between the two proteins. This makes the study of PleD, a GGDEF protein that does not have a role in cellulose regulation, and WspR, highly interesting.

The studies of native WspR and PleD proteins show that neither wild-type can substitute for the other in their respective organisms. This might be because they do not share inputs or because they do not have similar enough outputs. The use of a signal-independent WspR, which did function in *C. crescentus*, implied that the latter was the case. This was further confirmed by the fact that an allele containing the input domain of PleD and the output domain of WspR functioned in *C. crescentus,* but not in *P. fluorescens*. The crossover in output rather than input function is expected in terms of homology: the C termini of the two proteins show the highest similarity, and PleD differs form WspR in possessing two N termini.

The crossover did not, however, work in both directions. Neither signal-independent PleD, nor either hybrid of the WspR19 N terminus and the PleD C terminus, produced the WS phenotype in *P. fluorescens*. The reasons behind this are open to speculation, though the situation is not without precedent. Bock *et al.* (2001) used a more rational approach to the design of hybrids, using mass spectrometry of thermolysin-degraded fragments to identify the linker and the best place to make the break. However they found that the hybrid response regulator could fulfil only some of the functions of its output-domain parent. They suggest

that specific, inter-domain, interactions may be necessary for some functions and that these are disturbed in the hybrid. This is quite possibly the case with WspR, where a liberated C terminus was not functional. It may be that the requirements for the functioning of PleD are not quite as strict as those for WspR, hence the interchangeability only works in one direction. It is conceivable that PleD does not use cyclic di-GMP in its interaction with the flagellar motor, and it is instead a protein-protein interaction. Such a protein-protein interaction would be expected to be less constrained than an enzymatic one. If this is the case it would be very interesting from an evolutionary point of view, showing a change in mechanism between homologues.

Chapter 5

Fitness Consequences of *wspR* Alleles

The preceding two chapters have addressed mechanistic questions about *wspR*. In this chapter the study returns to the evolutionary environment that was the subject of the original genetic analysis (Kahn, 1998). *WspR* is examined as a candidate gene in the evolution of the WS adaptation. The fitness consequences of the alleles studied in Chapter 3 are analysed under a range of relevant conditions.

Aims:

1. To find out if *wspR* alleles are capable of increasing fitness of the ancestor
2. To discover whether LSWS is at a fitness optimum
3. To assess trade-offs between competitive abilities in two environments
4. To see if fitness varies with time and frequency
5. To confirm the behaviour of the phosphorylation-site mutants

5.1 Introduction

This study has treated *wspR* as a candidate gene for adaptive evolution. A genetic screen of an evolutionary transition (Rainey and Travisano, 1998) has identified it as being necessary for the biofilm-producing mutant, and homology and its position in the putative pathway suggest it may be important. The previous two chapters of this study provided further evidence that *wspR* is of potential importance in this evolutionary event because of the phenotypic effects of overexpressing mutant alleles. The candidate gene approach to studying evolution has been discussed at greater length in Chapter 1, along with its limitations. One major limitation in most studies carried out so far (eg Feder, 1999) is that they are forced to infer, rather than demonstrate, a link between phenotype and fitness. The manipulation of the gene in question often produces phenotypes of putative evolutionary significance such as the ones observed in this study; however, if no measurement of fitness change due to the altered phenotype can be made, there is no way of telling if it is a route that evolution can potentially take. In multicellular organisms fitness has to be estimated from a series of life-history trait measurements and these may prove unreliable. Many of the alterations caused to metazoans by gene manipulation may be detrimental to fitness despite their superficial resemblance to a known adaptation. Indeed the extreme position (Mayr and Provine, 1980) is that any mutation that is large enough to be visible is too large to be adaptive. The advantage of manipulating a bacterium such as *P. fluorescens* is that fitness can be measured directly on an experimental time-scale; this is made an even greater advantage in this case where the relevant system in which fitness is to be measured is a laboratory-based microcosm.

In this chapter the *wspR* mutant alleles that were analysed in Chapters 3 and 4 are given an evolutionary context. In those chapters they were assessed on the basis of various different phenotypes with the aim of understanding their mechanistic interaction with both the ancestral SM and derived LSWS genotypes. While this helped unravel the molecular details of WspR function it did not indicate whether any of the mutations in the gene sequence could

actually enhance fitness. An initial hypothesis might be that the dominant alleles, *wspR14* and *wspR19* would enhance the fitness of both SM and LSWS genotypes, as they increase cellulose production and hence biofilm formation. However, the other possibilities are that one or both of them produce effects that are too large to be adaptive when introduced into the ancestor, and that LSWS produces an optimal amount of cellulose so that any increase would reduce fitness. The C-terminal mutants might be expected to reduce the fitness of LSWS to the level of the ancestral SM by suppressing cellulose production. Whether or not this is the case will depend on there being complete complementation of the LSWS phenotype without other pleiotropic effects on fitness that these dominant-negative alleles cannot reverse. Allele *wspR*12, the wild-type, would be expected to have an intermediate effect on the ancestor. There may also be side effects caused by certain alleles that affect fitness in the opposite way to the phenotype that is obvious. Therefore in this chapter the different *wspR* allele-carrying genotypes are competed against controls in the microcosms from which the LSWS genotype was originally isolated. In order to look at pleiotropic effects of alleles in space and time, different conditions have been used for the competitions. To detect the presence of competitive trade-offs the competitions have been performed in both spatially structured and unstructured environments. It is expected that genotypes that increase fitness in one environment will decrease it in the other. The timescale of the competitions has also been varied to assess whether competitive ability depends on the growth phase. Finally, Rainey and Travisano (1998) found that frequency-dependent selection was responsible for maintaining genotypes in the microcosm population. Therefore, the starting frequencies of the competed genotypes are varied to assess whether this applies to the genotypes studied here.

Statistical significance was calculated in the first instance by the use of one-way or two-way ANOVA, as appropriate. Further individual comparisons, with respect to the relevant control, were performed using a two-tailed t test. In such cases the *P* values cited in the text have been adjusted with the Bonferroni correction.

5.2 Results

5.2.1 Fitness effects of *wspR* alleles on SM in a spatially structured microcosm

The initial fitness test performed was to see whether the *wspR* alleles used in Chapter 3 (the dominant-negative alleles *wspR5*, *wspR9*, and *wspR13*, the dominant alleles *wspR14*, and *wspR19*, and the wild-type allele, *wspR12*) enhanced the fitness of the ancestor in the static microcosm from which the WS phenotype was initially isolated. The genotypes overexpressing these alleles, along with one containing non-recombinant pVSP61 as a plasmid control, were competed against both a SM and a WS competitor in KB microcosms containing 0.0024% pantothenate. The SM competitor was a pantothenate mutant of the SM ancestor, which is selectively neutral in a pantothenate supplemented medium (P.B. Rainey, personal communication). The WS competitor was a pantothenate mutant WS genotype isolated independently from a microcosm. This is not identical to the LSWS genotype and can not be assumed to have the same fitness. Both pantothenate genotypes also contained pVSP61. Competitions were carried out for 24 hours as described in Chapter 2, and the results are shown in Figure 5.1 (top left panel).

5.2.2 Fitness effects of *wspR* alleles on LSWS in a spatially structured microcosm

To assess whether the changes caused to LSWS by the overexpression of *wspR* alleles correspond to fitness changes, the genotypes were competed against SM*panB* and WS*panB* in the same way as above, and the results are displayed in Figure 5.1 (top right panel).

5.2.3 Fitness effects of *wspR* alleles on SM in an unstructured microcosm

The enhanced fitness of WS genotypes in spatially structured microcosms is gained at the expense of a reduced fitness in unstructured microcosms (Rainey and Travisano, 1998). It is expected that similar trade-offs will be seen with the genotypes overexpressing *wspR* alleles.

Therefore the SM *wspR*-overexpressing genotypes were competed as in Section 5.2.1, except in a shaken microcosm and the results are shown in Figure 5.1 (bottom left panel).

5.2.4. Fitness effects of *wspR* alleles on LSWS in an unstructured microcosm

The fourth set of competitions shows the potential trade-offs in the effect of the alleles on LSWS. This was performed as in Section 5.2.2 above, except in a shaken environment and the results are shown in Figure 5.1 (bottom right panel).

5.2.5 Statistical analysis of 24 hour competitions

The data displayed in Figure 5.1 was treated as two sets: the effect of *wspR* alleles on SM in both environments, and the effect of *wspR* alleles on LSWS in both environments. In the former case, in competition with SM*panB*, there was found to be significant variation in fitness due to the environment (P=0.0002). There was also significant variation due to *wspR* allele (P=0.0002) and a significant interaction between allele and environment (P<0.0001). The alleles were divided into two classes: the dominant-negatives (*wspR5*, *wspR9* and *wspR13*), and those with a dominant effect on the ancestral SM (*wspR12*, *wspR14* and *wspR19*). In the structured environment there was found to be significant variation both within (P<0.0001) and between (P<0.0001) classes, though only the dominant class differs significantly from the control (P<0.05). In the unstructured environment there was significant variation between classes (P<0.0001), but not within classes (P=0.3664). Here again it is only the dominant class that differs significantly from the control (P<0.05). The equivalent data for the competitions against WS*panB* show the same trends, and the graphs illustrate that this is the case. These results are consistent with the dominant negative alleles having no effect on the ancestral SM. They are also consistent with the effect of the dominant alleles being allele-specific: *wspR12* and *wspR14* increase fitness in the structured environment, as expected, but *wspR19* diminishes it. The increased fitness conferred by *wspR12* and *wspR14* in the

structured environment is traded off against diminished fitness in the unstructured environment.

In the case of the *wspR* alleles overexpressed in LSWS, competed against SM*panB*, an identical two-way ANOVA showed significant variation due to *wspR* allele ($P<0.0001$) but not due to environment ($P=0.7172$), and a significant interaction between them ($P=0.0005$). When the alleles were divided into the same two classes as above (dominant and dominant-negative) there was shown to be variation between classes in both environments ($P<0.0001$), but never within classes. In the structured environment both classes diminished fitness significantly ($P<0.05$), while in the unstructured environment this was only the case with the dominant alleles. The equivalent data for the competitions against WS*panB* show the same trends, except that in this case there is significant variation due to environment ($P<0.0001$). These results are consistent with LSWS being at an adaptive peak, with any increase (due to the dominant alleles), or decrease (due to the dominant-negative alleles) in cellulose levels diminishing fitness in the structured environment. In the unstructured environment the dominant alleles diminish fitness as expected, while the dominant-negative alleles leave it unchanged. It might have been expected that the dominant-negative alleles would enhance fitness as the ancestral SM genotype is fitter than LSWS in this environment.

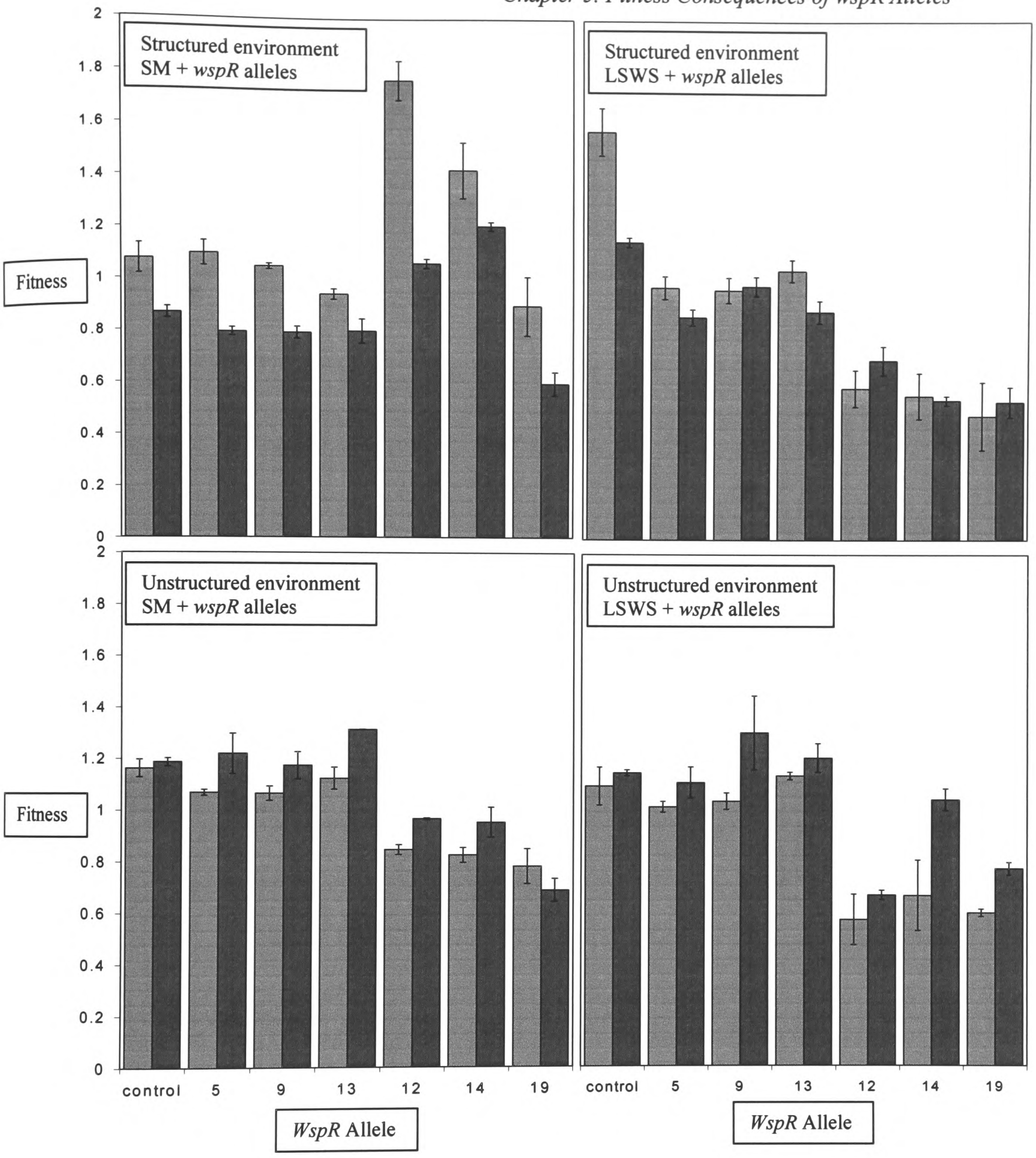


Figure 5.1. The effect of overexpressing *wspR* alleles on the fitness of SM and LSWS in 24 hour competitions. The fitness of SM and LSWS genotypes overexpressing 6 *wspR* alleles and a plasmid control was measured in competition with both SM*panB* and WS*panB* over 24 hours in static and shaken KB+pan environments. The light bars represent the competitons against SM*panB* and the dark bars against WS*panB*. In each case three replicates were performed and the error bars represent the standard error. The top two panels show the effects of the alleles on SM (top left) and LSWS (top right) in a structured environment while the bottom two panels show the effects of the alleles on SM (bottom left) and LSWS (bottom right) in an unstructured environment. In each graph the pVSP61 control is followed by the three dominant-negative alleles (*wspR5*, *wspR9* and *wspR13*), the wild-type allele (*wspR12*), and then the dominant alleles (*wspR14* and *wspR19*).

5.2.6 Fitness effects in 7-day competitions

The competitions performed above were all over a 24 hour period. It is possible that the fitness of certain genotypes depends on the timescale considered. For example a WS genotype grows more slowly than a SM genotype, but is able to colonise the air-broth interface better. Therefore the fitness of WS genotypes assessed before the end of log-phase might be lower than if assessed after several days of growth. To allow for some of these effects to become apparent certain genotypes were competed for 7 days in the structured microcosm. An added feature of this timescale is the possibility of mutations arising and being selected within either of the populations being studied. When a SM genotype is competed against a WS genotype, further WS genotypes could arise within the SM population, though it is unlikely that they will fare well unless they have a selective advantage over the existing WS genotypes. On the other hand, when two phenotypically similar genotypes are competed diversity is likely to arise within one or both to fill the available niches. The ability of the *wspR* overexpressing genotypes to diversify when appropriate and to outcompete potential diversification in a competitor are important facets of their fitness. The genotypes chosen for 7-day competitions were the two control genotypes [SM(pVSP61) and LSWS(pVSP61)], SM(pVSP61-*wspR9*) as a putative null result, SM(pVSP61-*wspR19*), SM(pVSP61-*wspR12*) and SM(pVSP61-*wspR14*) as the potential fitness-increasing alleles, LSWS(pVSP61-*wspR9*) as a representative dominant-negative, and LSWS(pVSP61-*wspR14*) as a fitness-diminishing overproducer of cellulose. All genotypes were competed against SM*panB* and WS*panB* and the results are shown in Figure 5.2. There was shown to be significant variation due to genotype by one-way ANOVA ($P<0.0001$ for both competitions). The very large fitness of SM(pVSP61-*wspR12*) versus SM*panB* is the most striking feature of the graph ($P=0.026$); SM(pVSP61-*wspR14*) also appears to have a large fitness advantage over SM*panB*, though this is not significant after the Bonferroni correction. Both genotypes also appear to outcompete WS*panB*, but this effect is not significant. Another feature is the fitness of SM(pVSP61-*wspR19*) when competed against SM*panB*; this contrasts with its low fitness over 24 hours. It is, however,

still unfit against WS*panB* (P=0.0055). Many of the other genotypes have fitness close to unity showing normal ability to diversify. The effect of *wspR14* on LSWS, however, allows it to compete well against SM*panB*, but not against WS*panB* (P=0.0001). This might be because it can colonise the air-broth interface better than SM*panB*, but not as well as WS*panB*. It was also observed, non-numerically, that SM(pVSP61-*wspR12*) and SM(pVSP61-*wspR14*) showed very little diversity at the end of their competitions, compared with that observed from all other populations, implying a degree of phenotypic plasticity. The full implications of these results will be considered in the discussion at the end of this chapter.

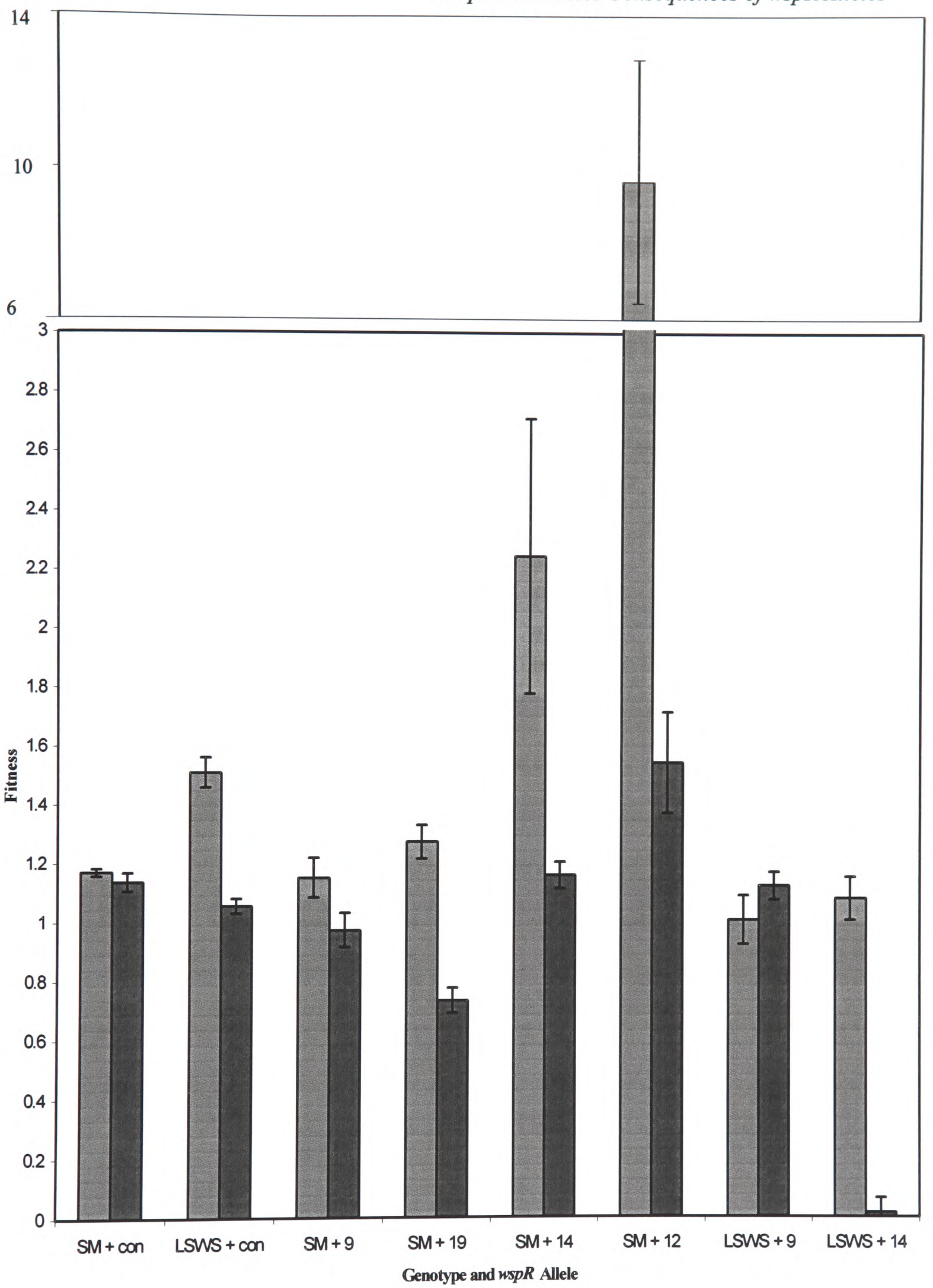


Figure 5.2. The fitness effects of overexpressing *wspR* alleles on SM and LSWS in a structured environment over 7 days. The fitness of SM and LSWS genotypes overexpressing different *wspR* alleles and plasmid controls was measured in competition with both SM*panB* and WS*panB* over 7 days in a static KB+pan environment. The light bars represent the competitons against SM*panB* and the dark bars against WS*panB*. In each case three replicates were performed and the error bars represent the standard error.

5.2.7. Frequency-dependent fitness

The competitions analysed thus far have been from 50:50 starting cultures. However, an important feature of the genotypes evolved from the structured environment is that they are subject to negative frequency-dependent selection. While the ancestral SM is less fit than the LSWS genotype when competed 1:1 it is still able to invade a population from rare. This is to be expected as there is a niche which the SM can colonise better than the WS, albeit a smaller one. To test for frequency-dependent effects several invasion-from-rare competitions were performed in spatially structured microcosms as described in Chapter 2. In all three sets of invasions described below *wspR9* was chosen to represent the C terminal mutants, while *wspR12*, *wspR14* and *wspR19* were all assessed individually.

5.2.7.1. Genotypes overexpressing *wspR* alleles invading SM*panB*

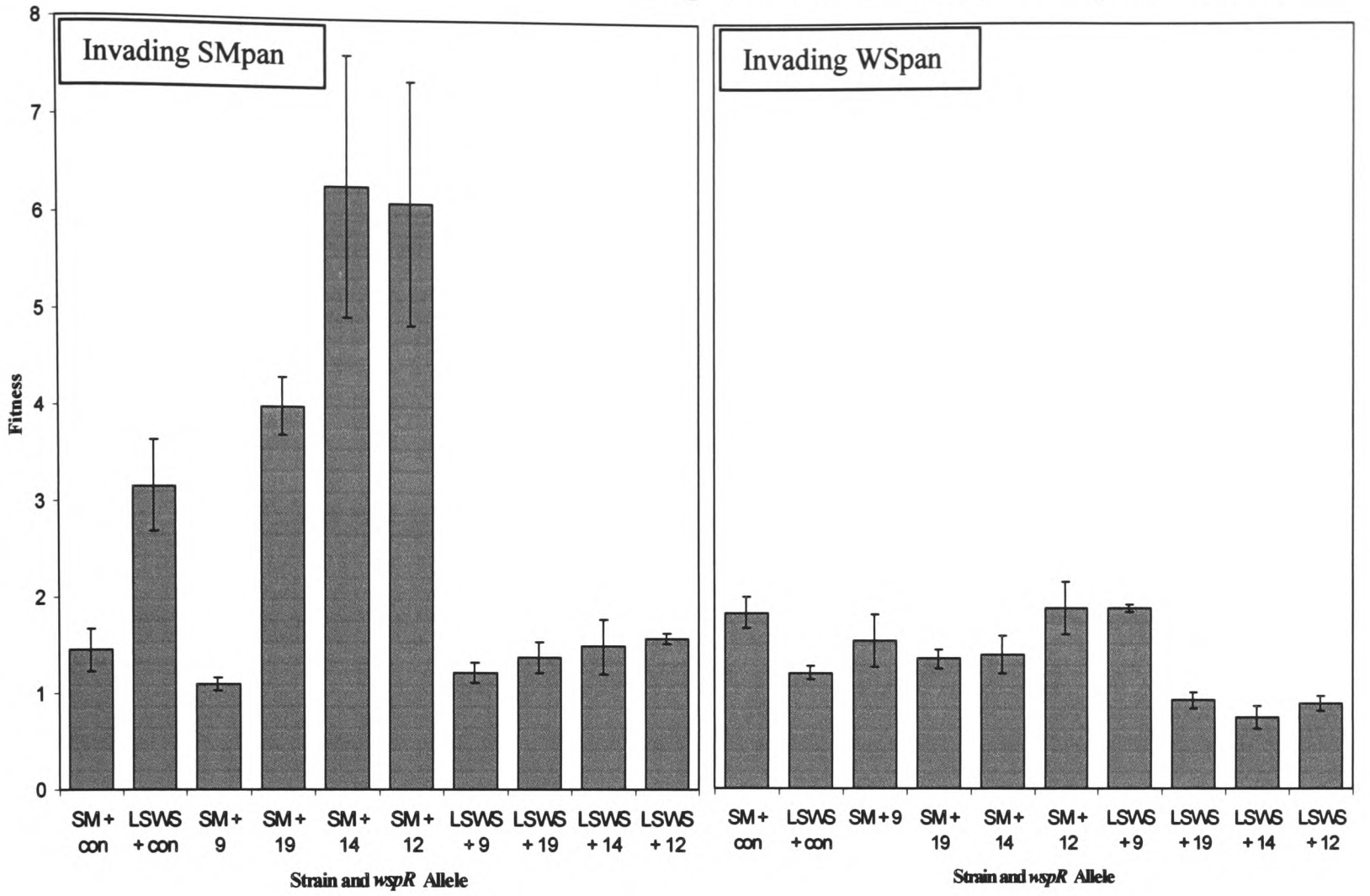
The results of 10 genotypes invading a population of SM*panB* from rare are shown in Figure 5.3 (top left panel). There was shown to be significant variation due to genotype by one-way ANOVA ($P<0.0001$). The fitness of the ancestral SM is slightly greater than expected (1.45, as opposed to unity). The fitnesses of both genotypes overexpressing *wspR9* and of all the modifications to LSWS are not significantly different from this showing no conclusive ability to invade from rare. The LSWS genotype appears to show the expected ability to invade though it is not significant after the Bonferroni correction. Neither are the even higher invasion fitnesses of SM(pVSP61-*wspR14*) and SM(pVSP61-*wspR12*). The only result that is significant is that of SM(pVSP61-*wspR19*) ($P=0.0117$); this is in contrast to this genotype's lower fitness in other assays. These results are generally as expected, with the WS genotypes invading SM*panB* and the SM genotypes failing to invade. The LSWS genotypes expressing dominant *wspR* alleles fail to invade.

5.2.7.2. Genotypes overexpressing *wspR* alleles invading WS*panB*

The results of the same 10 genotypes as above invading a population of WS*panB* from rare are shown in Figure 5.3 (top right panel). There was shown to be significant variation due to genotype by one-way ANOVA ($P<0.0001$). As expected the SM ancestor has a moderate ability to invade with a fitness of 1.83. SM(pVSP61-*wspR9*) and SM(pVSP61-*wspR14*) have fitnesses that may be as great as SM, but may be as little as LSWS (they are not significantly different from either). SM(pVSP61-*wspR19*) is more similar to LSWS and SM(pVSP61-*wspR12*) is more similar to SM, though these differences are not significant. LSWS+9 is also similar to SM and significantly greater than LSWS ($P=0.0058$), while the other alleles in LSWS lack the ability to invade from rare ($P<0.05$). These results are as expected with only the SM genotypes having the ability to invade a WS*panB* population.

5.2.7.3. The ability of SM*panB* to invade populations of *wspR*-overexpressing genotypes

While LSWS has the ability to invade SM, the converse is also true, with SM being able to invade. To test how this applies to the genotypes expressing *wspR* alleles, the ability of SM*panB* to invade them was measured and the results are displayed in Figure 5.3 (bottom panel). There was shown to be significant variation due to genotype by one-way ANOVA ($P=0.00162$). While the error bars are large and none of the individual fitnesses differ significantly from the controls, the tentative inferences are that SM, SM(pVSP61-*wspR9*) and LSWS(pVSP61-*wspR9*) (the phenotypically SM genotypes) cannot be invaded while the phenotypically WS genotypes can. The only exception is the surprising result of SM(pVSP61-*wspR19*) which seems resistant to invasion, though even this result is not significant.



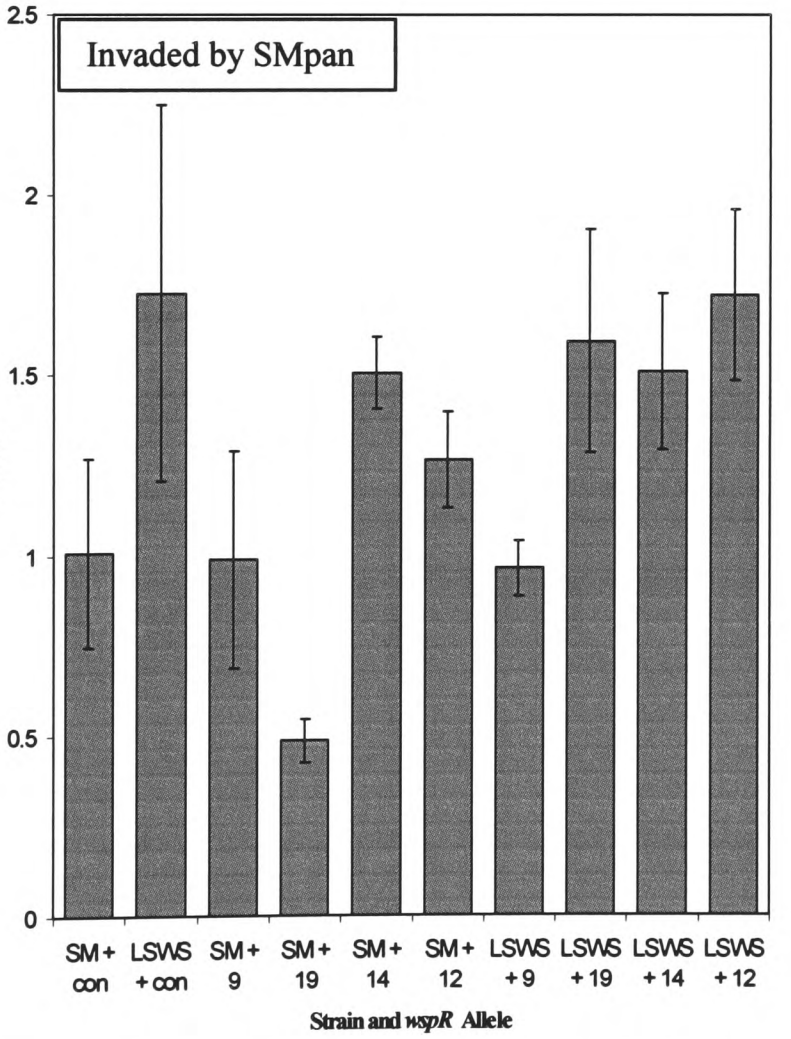


Figure 5.3. The fitness effects of overexpression of *wspR* alleles on SM and LSWS in a structured environment over 7 days when invading and invaded from rare. The fitness of SM and LSWS genotypes overexpressing different *wspR* alleles and plasmid controls was measured in competition with SM*panB* (top left panel) and WS*panB* (top right panel) over 7 days in a static KB+pan environment, invading from a frequency of 1000/1. In the bottom panel the fitness of SM*panB* is shown when invading populations of the *wspR*-overexpressing genotypes from a frequency of 1000/1. In each case three replicates were performed and the error bars represent the standard error.

5.2.8. The fitness consequences of *wspR* phosphorylation mutants

In Chapter 4 phosphorylation-deficient alleles of *wspR* were created, one based on the wild-type sequence and one on *wspR19*. These were cloned and overexpressed as with the alleles from Chapter 3 and their phenotypes scored. Several inferences about function were made from these results, so it is important to see if they behave like the other alleles in terms of fitness. It is expected that *wspR12d* (the phosphorylation mutant of the wild-type) will have no effect on fitness as its phenotype is completely null. On the other hand, *wspR19d* (the phosphorylation mutant of *wspR19*) is expected to alter fitness. If it behaves identically to *wspR19* it would be expected to be detrimental to the fitness of SM; however, the fact that it appears less severe in phenotype, might allow it to increase the fitness of the ancestor in the structured environment, in a similar manner to *wspR14*. Both alleles, expressed in both genotypes were competed against both SM*panB* and WS*panB*, in both structured and unstructured environments, in the standard 24 hour 50:50 assay.

5.2.8.1 Phosphorylation mutants in a spatially structured microcosm

The results of both genotypes (SM and LSWS), overexpressing both alleles (*wspR12d* and *wspR19d*) in a static microcosm are shown in the first panel of Figure 5.4; included in the figure is the data for both controls that has been shown in previous figures. There was shown to be significant variation due to allele by one-way ANOVA, performed separately for each combination of background and competitor (P=0.00522 for SM competed against SM*panB*, P<0.0001 for SM competed against WS*panB*, P=0.041 for LSWS competed against SM*panB*, and P=0.0002 for LSWS competed against WS*panB*). It can be seen that allele *wspR12d* (the wild-type with the aspartate to asparagine substitution) has no discernible effect on SM, but *wspR19d* (the same substitution in *wspR19*) increases the fitness of SM above that of SM*panB* and towards the level of LSWS (P=0.0018 in competition against WS*panB* and P=0.031 in competition against SM*panB*). This contrasts with the behaviour of the respective phosphorylation-proficient alleles: *wspR12* increases ancestral fitness while *wspR19*

diminishes it. The effect of *wspR12d* on LSWS is to diminish its fitness (not significant against SM*panB*, but P=0.0008 against WS*panB*) while *wspR19d* has a far less marked effect (not significant against SM*panB*; P=0.012 against WS *panB*). This fitness-diminishing effect of *wspR12d* is counter to the expectation that it would be completely neutral, and the small effect of *wspR19d* on LSWS is counter to the expectation that it should be detrimental.

5.2.8.2 Phosphorylation mutants in an unstructured microcosm

The second panel of Figure 5.4 shows the equivalent set of results to those in the first panel, in an unstructured microcosm. In three instances there was shown to be significant variation due to allele by one-way ANOVA (P=0.0483 for SM competed against SM*panB*, P=0.00183 for LSWS competed against SM*panB*, and P<0.0001 for LSWS competed against WS*panB*). The exception was the competition of SM genotypes against Ws*panB* (P=0.376). In the competitions against SM*panB* neither allele has a significant effect on the ancestor, though both appear detrimental (as expected for *wspR19d*). The effect of *wspR12d* on LSWS is not significant in either competition (though it appears detrimental in competition with SM*panB*) which is consistent with its purported lack of function; *wspR19d* has detrimental effects in both competitions (P=0.006 in competition against SM*panB* and P=0.0002 in competition with WS*panB*) as expected.

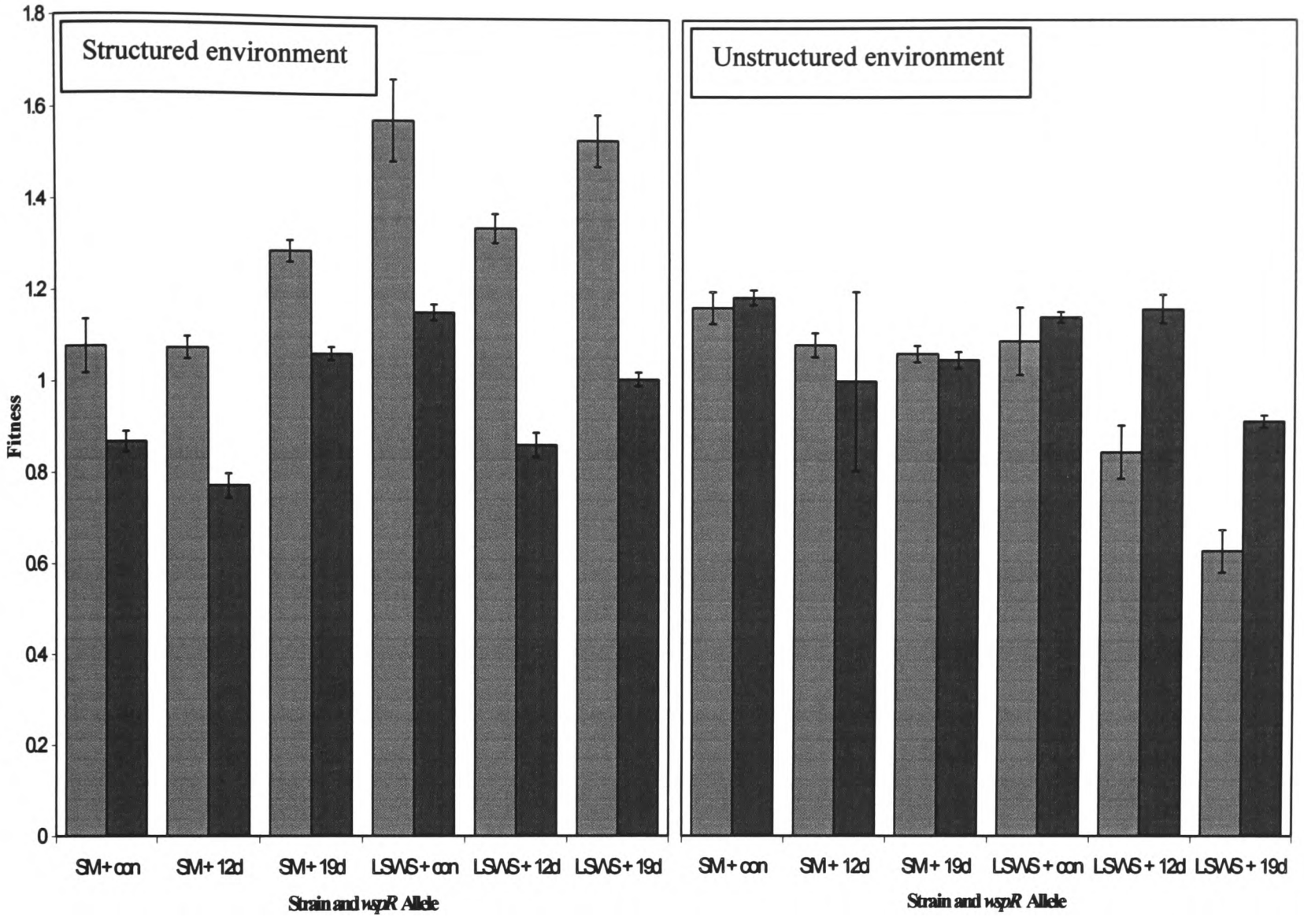


Figure 5.4. The fitness effects of overexpressing *wspR* phosphorylation mutants on SM and LSWS. The fitness of SM and LSWS genotypes overexpressing *wspR12d* and *wspR19d* and plasmid controls was measured in competition with both SM*panB* (light bars) and WS*panB* (dark bars) over 24 hours in both static (left hand panel) and shaken (right hand panel) KB+pan environments. In each case three replicates were performed and the error bars represent the standard error.

5.3 Discussion

5.3.1. Manipulating the fitness of the ancestor in a structured environment

The first set of data, presented in Section 5.2.1 and the first panel of Figure 5.1, looks at the use of *wspR* alleles to enhance the fitness of the ancestor. If *wspR* is an important gene in the production of the WS phenotype then evolution may have altered its sequence or expression levels to arrive at that phenotype. If so, it may be possible to recreate that process using artificial manipulation of *wspR* sequence, and plasmid-based overexpression. The earlier chapters of this study have already shown how the WS phenotype can be mimicked in this way; here the equivalent change in fitness is looked for. Do the *wspR* alleles that change a SM genotype into WS also enhance fitness? This first set of data therefore looks at this important question, as it looks at the fitness of manipulated SM genotypes in the structured environment from which LSWS was isolated.

The data in the first panel of Figure 5.1 show competitions against SM*panB* and WS*panB* following similar trends, with all the fitnesses against the latter being lower because of its enhanced fitness in this environment. The dominant negative alleles, *wspR9*, *wspR13* and *wspR5*, show little difference from the control in either competition, implying that the overexpression of a *wspR* allele is itself not costly enough to affect fitness. (The carriage of the plasmid may be costly, but the control also carries it so this experiment will not detect it). *WspR13* lowers the fitness of SM slightly, though little can be inferred from this. It is the allele that has two mutations from the wild-type sequence, but there is no way of telling from this data if that fact is relevant.

The alleles that have a dominant effect on the SM genotype, i.e. *wspR19*, *wspR14* and *wspR12*, have more striking effects on fitness. These dominant alleles, treated as a class, have a significant effect on fitness compared with the plasmid control. Both *wspR12* (the wild-type

allele) and *wspR14* increase the fitness of the ancestor, justifying the choice of *wspR* as a candidate gene. The difference between these two alleles is that *wspR14* is the fitter allele when competed against WS*panB*, whereas *wspR12* is the better allele (and the fittest overall) when competed against SM*panB*. This might have something to do with SM(pVSP61-*wspR12*) being more of a plastic generalist than the other genotypes and will be discussed more fully below.

Allele *wspR19*, on the contrary, decreases the fitness of the ancestor. This effect is particularly noticeable in the competition against WS*panB*, the large error bar in the competition against SM*panB* (and consequent lack of significance) making it harder to analyse. It may be that the detrimental side-effects of *wspR19* are greater than the positive effects or simply that it causes too much overproduction of cellulose. The latter view is supported by the fact that the assays in Chapter 3 imply that *wspR19* is the most severe allele.

5.3.2. LSWS and an adaptive peak

As LSWS was isolated from a microcosm evolved from an ancestor it might be expected to reside on an adaptive peak in terms of the extent of the WS phenotype. In other words it should have optimal levels of cellulose production and attachment so that perturbing the system should only result in decreases in fitness. The top right panel of Figure 5.1 shows that this is indeed the case; the highest levels of fitness are in the control competitions and both classes of allele (dominant and dominant-negative) decrease fitness. It should be noted from the controls that LSWS has a fitness greater than unity when competed against WS*panB*, emphasising that they cannot be considered identical genotypes.

The dominant-negative alleles reduce the fitness of LSWS back down to the level of the ancestor. This is exactly as would be expected from their phenotype. The dominant alleles, on the other hand, which result in WS phenotypes when overexpressed in LSWS, show fitnesses

even lower than the ancestor, in competitions against both SM*panB* and WS*panB*. It is thought that they are producing too much cellulose, taking them down the adaptive peak. Their lack of ability to compete with WS*panB* can be explained by the fact that WS*panB* colonises their shared niche better. The low fitness against SM*panB* shows that they are bad colonisers of their niche even in the absence of direct competition, at least in the timescale measured.

5.3.3. Trade-offs in unstructured microcosms

As mentioned in Section 5.2.3 the fitness advantage of LSWS is gained at the expense of its ability to compete in an unstructured environment. It is therefore expected that the artificially created genotypes, which have fitnesses akin to LSWS in the structured environment, will have the same trade-off. In the bottom left panel of Figure 5.1 this is seen to be the case with *wspR12*, *wspR14* and *wspR19* reducing the fitness of the ancestor in such an environment. The two-way ANOVAs of competitions involving the *wspR* alleles in the ancestral SM showed significant effect of environment as well as a significant interaction between allele and environment. This is consistent with there being a trade-off. This experiment is in some ways the reverse of that performed in Section 5.2.2, as in this case the ancestor can be expected to be at a fitness peak and alterations should diminish it. This is not quite the case, though, as the dominant-negative alleles do not diminish fitness. This is consistent, however, with their apparent lack of phenotypic effect. The possible exception is *wspR13* which might be having a positive effect on ancestral fitness, albeit small. This is the same allele that diminished fitness in the structured environment so possibly we are seeing the opposite trade-off. It is unclear what the mechanistic explanation for this might be.

The effects of *wspR* alleles on the fitness of LSWS, as shown in Figure 5.1, bottom right panel, are harder to interpret. The main problem is that the fitness of the LSWS control is close to unity when competed against the ancestor even though it was expected to be lower

because there is trade-off between its enhanced fitness in a structured environment, and its fitness in an unstructured environment. However, if it is assumed that this result is unnaturally high the other results make sense. The dominant-negative alleles give LSWS the same fitness as the ancestor while the dominant alleles decrease it further. The two-way ANOVAs of the competitions involving *wspR* alleles in LSWS show no significant variation due to environment. This is because the dominant alleles decrease fitness in all environments (consistent with LSWS being at an adaptive peak) and the dominant-negative alleles cause a reversion to the SM phenotype in all cases. The other unexpected result is that *wspR14*, in contrast with *wspR12* and *wspR19*, is relatively fit against WS*panB*. There may be something in the way in which WspR14 acts that makes it no less fit than another WS genotype. This result, and the generalist phenotype of SM(pVSP61-*wspR12*) discussed below, raises questions about the universality of trade-offs, that have been considered by other authors (e.g. Fry, 1996).

The general conclusion form the short-term competitions is that *wspR* alleles that produce similar phenotypes may have different effects on fitness. This is not particularly apparent amongst the dominant-negative alleles which have the predicted effects. However, even within this class, there is the possible exception of *wspR13* which might be making the ancestor more SM-like. The differences are most apparent in the dominant alleles. In general *wspR12* and *wspR14* behave similarly, with *wspR19* acting differently. This is most pronounced in the effects on the ancestor, with *wspR12* and *wspR14* increasing its fitness in a static microcosm, but *wspR19* diminishing it. There are more subtle differences between *wspR12* and *wspR14* that have been discussed above. It might be expected, from the sequences of the alleles, that *wspR14* and *wspR19* would be more similar and that *wspR12*, being the wild-type allele, behaves differently. Another counter-intuitive feature is the low fitness of LSWS with these alleles, though this is explained by the consideration of adaptive peaks. These results make further investigation over longer timescales all the more pertinent.

5.3.4. Long-term competitions

Long term competitions allow a wider range of factors to affect fitness. Two of these factors are competition after log phase, which is more likely to involve K-selection, and diversification, which is especially likely when the two genotypes being competed naturally tend to colonise the same niche, leaving other niches available. This is less of a factor when a SM and a WS genotype are competed though there is also the niche normally colonised by the fuzzy spreader (FS). The long-term experiments were only carried out in static microcosms for a variety of reasons, including the fact that diversifying selection does not act in an unstructured environment (Rainey and Travisano, 1998).

The most striking feature of Figure 5.2 is the very large fitness of SM(pVSP61-*wspR12*) when competed against SM*panB*. This level of fitness means that the genotype almost eradicates any competition from other genotypes. The fact that very low diversity in colony morphology is found in the descendants of SM(pVSP61-*wspR12*), coupled with its large fitness, imply that it can colonise several environmental niches without the need for genetic diversification. If this is the case (it is possible that there is diversification that has no effect on morphology) then the genotype is a phenotypically-plastic generalist that is able to switch between the abilities to colonise different niches. This is consistent with the conditional phenotype observed on plates that depends on salt. In terms of the model proposed in Chapter 3 this genotype may have heightened sensitivity to some environmental signal that allows it to behave differently in each niche. Such an argument is consistent with the role that several authors have proposed for regulatory genes in the evolution of plasticity (Schlichting and Pigliucci, 1993; Via, 1993a; Scheiner, 1993) and the tendency for plasticity to evolve in a heterogeneous environment (Gregorius, 2000; De Jong, 1995; Via, 1993b). It may be that SM(pVSP61-*wspR12*) can grow as fast as the ancestor in log phase before turning on cellulose production and consequent surface niche colonisation. However, in competition with WS*panB*, SM(pVSP61-*wspR12*) performs well, but not to the extent that it does against

SM*panB*; this is to be expected if WS*panB* can colonise the largest niche as quickly as SM(pVSP61-*wspR12*) can, but is not entirely consistent with the previous theory, as in that SM(pVSP61-*wspR12*) would be expected to grow faster during log phase and then colonise the air-broth interface before WS*panB*. It may be that the plasticity is not perfect and that there is some advantage to being a specialist.

Far less striking than SM(pVSP61-*wspR12*) is the fitness of SM(pVSP61-*wspR14*). However this is still very fit and has a larger fitness over 7 days than in equivalent competitions over 24 hours. More notable is the reversal in the fate of SM(pVSP61-*wspR19*). In 24 hour competitions this did not compete well with SM*panB* and it was thought that it produced too much cellulose. In the longer term it does compete well so it might be that the overproduction of cellulose causes it to grow slowly but it can successfully occupy the surface niche once it has reached stationary phase. Presumably it is able to do this before the SM genotype has diversified to fill the niche itself.

It is notable that SM(pVSP61-*wspR12*) both outcompetes WS*panB* and, more importantly, does better in both competitions than the LSWS control. That means that, for these particular environmental conditions at least, a fitter genotype than the one produced by natural selection (LSWS) has been created. This does assume that LSWS was the fittest genotype isolated from the initial microcosms or was at least sufficiently representative that SM(pVSP61-*wspR12*) is fitter than all of them. If this is the case this, then it is an example of a route that was not open to selection, either because the necessary mutations did not occur or because the intervening genotypes were disadvantageous. The WS genotypes isolated from microcosms may represent local adaptive peaks whereas this artificially-engineered genotype is a higher but inaccessible peak. The reasons behind the existence of these fitness valleys can only be speculated on, but if they are real this experiment has demonstrated the existence of a constraint. Further discussion of the nature of this constraint will be made in Chapter 6.

The genotypes expressing dominant-negative alleles do not behave significantly differently from the SM control in 7 day experiments (Figure 5.2.); this is to be expected if the effect on SM is neutral and the effect on LSWS is to reverse the WS phenotype completely. The addition of *wspR14* to LSWS produces a very unfit genotype in 24 hour competitions. This is again the case with the long-term competition against WS*panB*, which is to be expected if the optimum amount of cellulose production has been exceeded. This genotype is however able to compete against SM*panB*. This implies that, though slow-growing, it can still fill the surface niche before the necessary diversification has occurred in its competitor. This situation is similar to that observed with SM(pVSP61-*wspR19*).

5.3.5. Frequency-dependent selection

Unlike the other competitions shown above, the numerical values for fitness are of less importance when considering invasions from rare. The important factor here is the ability or lack of it to invade a population; the ability to do so showing a negative frequency-dependent selective advantage. Hence fitness values close to and below unity show no evidence of invasive ability, whereas those significantly greater than unity do. The first panel of Figure 5.3 illustrates this point well.

This data set shows that LSWS, and the ancestral SM genotypes that have been modified into WS genotypes, have the ability to invade a SM population from rare. This is to be expected as there is a niche which they can colonise far better than the competitor. The genotypes that are phenotypically SM are not able to invade the population as they have no particular advantage in colonising any niche against other SM genotypes. The phenotypically WS, modified LSWS genotypes, which are very unfit in the competitions performed above, show fitnesses slightly greater than one. It is hard to assess whether this really constitutes an invasive ability. While there is a niche available to them they may be so unfit due to the excess cellulose, that the SM*panB* and its descendants are better able to colonise the air-broth interface. This would

depend on the precise balance of the relative fitnesses and the rate of mutation to the WS phenotype. It should be noted in these analyses that it is not necessarily correct to assume that niches have a fixed physical size and population capacity. The capacity of one niche might be governed by the extent of occupation of another, or by the precise nature of the clone occupying it. An obvious example is the niche occupied by the ancestor having its air and nutrient supply reduced by the formation of a biofilm, and there are likely to be more subtle cases as well.

The ability of genotypes to invade a population of WS*panB*, as illustrated in the second panel of Figure 5.3, is less clear cut due to the larger errors and smaller differences between genotypes. As expected, the phenotypically WS, modified LSWS genotypes are unable to invade as they have no niche open to them that is not open to a competitor and suffer the other diminutions in fitness discussed previously. The other genotypes appear to fall into two categories, though the errors make the distinction less than rigorous. The phenotypically WS genotypes, LSWS, SM(pVSP61-*wspR19*) and SM(pVSP61-*wspR14*) have some ability to invade WS*panB* from rare. These are the genotypes that are competing directly with WS*panB* for a niche so the basis of their ability is unclear. The phenotypically SM genotypes, SM, SM(pVSP61-*wspR9*) and LSWS(pVSP61-*wspR9*) have higher invasive abilities which can be interpreted in terms of their colonisation of a different niche. Interestingly SM(pVSP61-*wspR12*) also appears to be in this category which is consistent with the suggestion made above that it is phenotypically plastic and can colonise different niches.

The final set of invasion experiments, shown in the third panel of Figure 5.3, looks at the ability of genotypes to resist invasions from the ancestral SM*panB*. This is pertinent in showing whether genotypes have superseded the fitness of their ancestor in the same niche, or diverged from it to live in parallel in a different niche. For this reason, as well as the fact that no LSWS*panB* genotype was available, susceptibility to invasion by a WS genotype was not looked at.

As mentioned already the errors in this data set are very large and none of the results are significantly different from the controls. It is however clear that the phenotypically SM genotypes cannot be invaded, which is as expected. With one exception, it appears that the other genotypes can, which is also to be expected because of they occupy a different niche from the ancestor. It might be important that SM(pVSP61-*wspR12*), the putative plastic genotype, has the lowest susceptibility, maybe because it is already occupying the ancestor's niche. The anomalous result of SM(pVSP61-*wspR19*) is very hard to explain. If the result is genuine and not merely an artefact, it might be because of some serious effect that the genotype has on the whole environment which affects colonisation of the other niches. The other assay that showed this genotype to be unique was the study of *wss* transcription in Chapter 3 in which not even LSWS(pVSP61-*wspR19*) differed from the other genotypes. Whether this is relevant remains unclear.

5.3.6. Phosphorylation mutants

These mutants were created in Chapter 4 to assess the role of phosphorylation in *wspR* function. They were assayed phenotypically and certain inferences were made about them. It is of great interest to find out whether the phenotypes correspond to fitness as expected, because this would add extra validity to the conclusions that were derived in that chapter.

The wild-type phosphorylation mutant, *wspR12d*, was thought to be completely neutral as its function is thought to depend on the input from a histidine kinase. When added to the ancestor this appears to be the case, in both environments, though it might have a slightly deleterious effect in the shaken one. The situation is, however, less intuitive when the allele is added to LSWS. In the structured environment the fitness of LSWS is reduced, to the level of SM, when in competition with WS*panB*, and to an intermediate level when in competition with SM*panB*. In Figure 5.2 it was seen that all alleles caused some decrease in the fitness of

LSWS, though the magnitude differed between dominant-negative and dominant alleles, the latter having a greater effect. Hence there are at least two possibilities for the reason behind the effect caused by allele *wspR19d*. The magnitude of the fitness decrease is most similar to that seen with dominant negative alleles. In Chapter 4 the failure of this allele to cause a dominant-negative effect was cited as evidence that such an effect was due to the soaking up of the phosphorylation signal rather than the binding of WspR to the kinase. If this reduction in LSWS fitness seen here is because the allele is having some kind of dominant-negative effect then that conclusion needs reassessment. The intermediate fitness against SM*panB* may indicate that a combination of the two factors (soaking up of phosphate signal and kinase binding) is the real reason behind dominant negativity. The situation gets more complicated when the effect in a structured environment is considered. If the allele is having a dominant-negative effect then it might be expected to increase the fitness here. Instead it appears to do the opposite, against SM*panB* at least. However it has been suggested above that the fitness of the control is erroneously high in which case the allele would be having no effect on LSWS. The combination of these two results implies that there is more than one component to the phenomena of dominant-negativity and WspR function, the soaking up of the signal being the more important, but other factors being sufficient to hinder the fitness of LSWS in its optimum environment. On the other hand it may be that the *wspR12d* has a fitness effect that is just coincidentally similar to the dominant-negative alleles. It may not have lost function completely so is having a detrimental effect that is less severe than the other dominant alleles, so does not diminish LSWS fitness as much.

If *wspR19* is truly independent of an input signal, and if the input signal in question is phosphorylation at the aspartate residue, then *wspR19d* should behave in an identical way. This has already been shown in Chapter 4 not to be strictly true as the phenotype is less severe. The reasons for this have already been discussed and the allele may well still be signal-independent. In terms of fitness, the most striking effect is that in the structured environment, *wspR19d* increases the fitness of the ancestor where *wspR19* decreased it. In this

respect *wspR19d* behaves like *wspR12* and *wspR14*, which are less severe than *wspR19*. This is consistent with the phenotypic observations made in Chapter 4. It also diminishes the fitness of the ancestor in the unstructured environment, though perhaps not as much as might be expected. The effect of this allele on LSWS is to diminish fitness in the same way as do all dominant alleles. The only exception is in the static environment, against SM*panB*, where it has very little effect. The allele may be sufficiently mild that it does not hinder the ability to colonise an empty niche though in the assay against WS*panB*, where the competition for the niche is direct, it fares less well.

The fitness assays performed in this chapter have added an extra dimension to all the other phenotypic assays performed in Chapter 3. They have shown that *wspR* is a justified subject for a candidate gene study and have given insights into the LSWS phenotype in general. These include information about plasticity, adaptive peaks, constraints and trade-offs. Interestingly the study also gave some information about mechanistic questions such as dominant-negativity.

Chapter 6

WspR Mutants in Natural Populations

The phenotypic and fitness changes that are the result of the artificial manipulation of *wspR* suggest that *wspR* sequence variation may explain some of the naturally occurring WS genotypes. To test this the sequence of *wspR* was determined from WS genotypes isolated from independent microcosms. One of the artificially created genotypes [SM(pVSP61-*wspR12*)] was found, in Chapter 5, to be phenotypically plastic. Though this genotype was extremely fit there is no evidence that it occurs naturally; the reason for this is discussed in this chapter.

Aims:

1. To sequence *wspR* from 10 independent WS genotypes
2. To look for plastic genotypes in microcosm isolates
3. To evolve plastic genotypes artificially by removing constraints

6.1 Introduction

It had already been established (Kahn, 1998) that the mutation responsible for the LSWS phenotype was not in *wspR*. However, the preceding chapters of this study have shown that it is a candidate locus for such a mutation and that artificial manipulation of *wspR* can cause the transition from ancestral SM to evolved WS. In Chapter 3 the random mutations of alleles *wspR14* and *wspR19* were shown to have a dominant effect on the ancestral SM genotype when overexpressed. This contrasts with the conditional dominant phenotype caused by overexpression alone (the case of wild-type allele *wspR12*). What remains unclear is whether the mutations of alleles *wspR14* and *wspR19* alone, without an increase in copy number, are sufficient to cause the SM to WS transition. In other words does a chromosomal mutation in *wspR* in a SM background carrying no plasmid create a WS genotype? Such a mutation would need to cause the phenotypic characteristics of a WS genotype, such as CLP production and biofilm formation, and also the necessary increase in fitness. In Chapter 5 it was demonstrated that, in the case of allele *wspR19*, it was possible to create a WS morph that has reduced fitness (though this varies with the environmental conditions). Such a genotype would not arise by natural selection. Therefore it is important to know whether a mutation in *wspR* is found in any WS genotypes that have evolved in microcosms, or whether such mutations do not have the requisite effects on phenotype and fitness, or are outnumbered by other possible mutations. To address this question a set of independently isolated WS genotypes was required. Twenty six such genotypes were isolated (S. Kahn and P.B. Rainey, unpublished data), each from an independent microcosm, with the aim of sampling the entire range of phenotypic variation within the WS class. Hence from some microcosms common colony morphologies were selected whereas from other microcosms the selected morphology was deliberately rare. The set of WS genotypes were named WSa to WSz.

In Chapter 5 the high fitness of the ancestral genotype overexpressing wild-type *wspR* was attributed to phenotypic plasticity. If such a high fitness can be achieved by overexpression

alone it is interesting to know whether such genotypes arise naturally in microcosms and, if not, why not. There may be ecological reasons that do not favour plasticity; this seems unlikely at first given the high fitness of the plastic genotype in the structured environment when competed against the ancestor for seven days. These conditions most closely mimic those when the adaptive radiation itself occurs, but there may be other unknown factors. Alternatively, as suggested before, there may be a mechanistic constraint that prevents the evolution of plasticity in this case. This can be examined by assessing the phenotypes of the 26 WS genotypes with respect to salt-dependence (thought to indicate plasticity) and by finding circumstances in which plasticity can evolve.

6.2 Results

6.2.1 The sequence of *wspR* from independently isolated WS genotypes

From the 26 WS genotypes 10 were selected at random for sequencing analysis. This was done for practical reasons, while still having sufficient representation to assess the importance of *wspR* mutations in this adaptive radiation. In order to perform the sequencing *wspR* was amplified by PCR, using *Taq* polymerase and primers *wspR*-for and *wspR*-rev, from each selected genotype. This product was sequenced using primers *wspR*-S1 to *wspR*-S6 as done for the *wspR* alleles in Chapter 3, and the sequences were aligned and any ambiguities checked. All ten genotypes were found to have the wild-type sequence of *wspR*. While this does not preclude *wspR* being a possible site of mutation it indicates that, if it is, it accounts for only a minority of cases.

6.2.2. The evolution of plasticity

To test whether any of the 26 WS genotypes demonstrated the phenotypic plasticity observed when *wspR* is overexpressed in the ancestral SM genotype, their colony morphologies were compared on LB and LBns. No difference was found implying that, like LSWS, they were not phenotypically plastic. To test if this was due to a mechanistic constraint that prevented *wspR* overexpression arising, rather than ecological factors, all the *wspR* overexpressing genotypes (*wspR9*, *wspR19*, *wspR13*, *wspR14*, *wspR5* and *wspR12* in both SM and LSWS genotypes) and plasmid controls were evolved for 7 days in static microcosms containing kanamycin to maintain the plasmid. Each microcosm was plated out onto LB and LBns and in every case both SM and WS colonies were present (though fuzzy spreaders only evolved from the SM control). To test for plasticity SM colonies from the LBns plates were re-streaked onto LB and WS colonies from the LB plates were re-streaked onto LBns. In each case, with the exception of the controls, some colonies were found that were SM on LBns and WS on LB. The experiment was not performed in sufficiently large numbers to achieve robust numerical data regarding the frequency of plasticity, but it did demonstrate that plasticity evolves from

genotypes in which *wspR* is already overexpressed, whereas it does not evolve from the controls. The SM control is equivalent to the initial adaptive radiation.

6.3 Discussion

6.3.1 The absence of mutations in *wspR*

The ten randomly selected WS genotypes were all shown to have the same, wild-type, sequence of *wspR*. While this does not preclude mutations in *wspR* causing the WS phenotype it does indicate that their frequency is low and cannot be responsible for the majority of cases. The *wspR* mutations studied in Chapter 3 are capable of causing the transition when overexpressed and, in the case of *wspR14*, cause the requisite increase in fitness as well. It may be that overexpression is necessary for the effect, at least in alleles that are not so severe that they diminish fitness, and this would require a further mutation in a regulatory element. However, if this explains the fact that a mutation like that found in *wspR14* is not found in natural populations, it also suggests that an allele as strong as *wspR19* might have a fitness-enhancing effect when present on the chromosome and not overexpressed. Alternatively, *wspR19* is too strong even in single copy, and *wspR14* is weak enough to require secondary mutations in regulatory regions and for some unknown reason mutations of intermediate strength do not exist. It may be that there are so many other possible sites for mutations that one in *wspR* is unlikely to occur in a sample of ten. This may be merely the effect of numbers among equally probable mutations, as there are seven genes in the operon as well as regulatory elements and other important genes in other operons. Alternatively there could be a reason why mutations earlier in the pathway are more likely to increase fitness in the relevant manner than those in *wspR*, possibly because there is more scope for modification. The answer to this question may depend on the precise map of inter-related pathways and the extent to which different mutations have undesirable pleiotropic effects. It may also depend on the relevant likelihood of gain and loss-of-function mutations. Though the terminology is ambiguous and depends on the level studied, the mutations sought in *wspR* are essentially gain-of-function (even if they involve the loss of function of the N terminus). It may be that a loss-of-function mutation in another gene is more likely to occur because the number of ways of knocking out a gene far exceeds the number of ways of increasing its activity. In this case

one of the negative elements in the *wsp* operon, such as *wspF,* which encodes the methylesterase, might be more likely candidates.

6.3.2 The evolution of plasticity

A related question is why no plastic mutant akin to SM(pVSP61-*wspR12*) evolves by the up-regulation of *wspR*. In Chapter 5 this genotype was shown to have a very high fitness in all assays, especially when competed against the ancestor in a 7 day experiment, but also when competed against LSWS. The checking for salt-dependence of the 26 WS genotypes confirms that the evolution of such plasticity does not occur despite the fitness advantages. If there were ecological reasons for this fact then plastic mutants would not be expected to evolve from any SM or WS genotypes. However, if there are mechanistic constraints, the precise genotype of the ancestor would be more important. The demonstration in this chapter that plastic mutants can arise from genotypes that are already overexpressing *wspR* alleles supports this latter view, especially as some of the genotypes in question have fitnesses that are equivalent to or higher than the ancestor. If the constraint in question is that it is impossible to increase *wspR* expression levels without also affecting those of the entire operon, as it is co-transcribed, then the presence of an expression plasmid containing only *wspR* should circumvent it. It may be that the wild-type copy of *wspR* present in the chromosome rescues the plasmid copy by homologous recombination, so that a genotype overexpressing the wild-type allele, rather than a mutant, is created. It is interesting that plasticity also arises in the LSWS genotypes overexpressing *wspR* alleles as this would require recombinational repair and a reversion in the chromosome. Whatever the precise mechanism, these results add weight to the suggestion that there is a constraint on the evolution of *wspR*-based plasticity.

Chapter 7

Further Analysis of the WS Phenotype

In this chapter the WS phenotype is studied more globally than before. A proteomic approach is used to identify differences between the ancestor and LSWS. As LSWS is only one of the many WS phenotypes that occur, and in terms of the mutation causing it, may be unique, other WS genotypes were profiled. This was done in a range of phenotypic assays which also assessed the role that *wspR* may or may not play in these genotypes.

Aims:

1. To determine the reason that LSWS is impaired in motility
2. To compare the proteomes of SM and LSWS
3. To characterise the phenotypes and fitnesses of 6 independently isolated WS genotypes
4. To assess the effect of *wspR* alleles on these genotypes

7.1 Introduction

So far this study has concentrated on *wspR*, which is one important component of the LSWS phenotype. In Chapter 4 its relationship with the *wsp* and *wss* operons was confirmed by the analysis of some epistatic interactions. However, it is very likely that WspR interacts with other factors, both upstream and downstream, the analysis of which might provide further insight into the LSWS phenotype. A related issue is the role that the *wsp* and *wss* operons in general, and *wspR* in particular, play in WS genotypes other than LSWS. It is unlikely that other genotypes contain the same mutation as LSWS, though many of them may have mutations in the same gene and even more in the same operon or pathway. However, there may be other routes to the WS phenotype that do not involve the *wsp* operon, or even potentially the *wss* operon, though this latter case is far less likely.

A potential additional downstream target of WspR is motility. This is suggested by several different lines of evidence. Firstly the functional homology between WspR and PleD (demonstrated in Chapter 4) implies that the former is capable of interacting with the flagella motor to cause its degradation, as the latter does in *C. crescentus* (Aldridge and Jenal, 1999). If this was the case then a *wspR* mutant may be supermotile because of the lack of negative regulation and a *wspR*-overexpressing genotype might be non-motile. Secondly, there is evidence, reviewed in Chapter 1, that motility plays an important role in biofilm initiation. This means that motility is required for attachment (Pratt and Kolter, 1998), but also that some biofilm-forming genotypes become non-motile because of hyperflagellation (Deziel *et al.*, 2001). For this reason, it would be interesting to know if *wspR* has an effect on motility and if so, whether it is positive or negative.

Most of this study so far has used a genetic approach to the study of *wspR* and the LSWS phenotype. While this approach is very powerful, and represents the level at which

evolutionary change occurs, the actual mechanisms occur at the level of proteins and their interactions. In the post-genomic era two-dimensional protein gel electrophoresis has been used for the comparative analysis of entire proteomes. This approach can profitably be applied to evolutionary problems such as the adaptive radiation of *P. fluorescens*, where two genotypes differ by as little as a single nucleotide in the DNA sequence but their phenotypes, and hence presumably their proteomes, differ far more. This is all the more useful in a system where it seems that regulation is occurring at a post-translational level (discussed in Chapter 3). An alternative technique, the DNA microarray, uses cDNA-mRNA hybridisation to look at global changes in mRNA expression levels. However this does not give information on processes that are regulated at the level of protein activity, and also it has been demonstrated that the correlation between mRNA and protein levels is weak (Gygi *et al.*, 1999).

The basic technology of proteomics is reviewed by Dutt and Lee (2000). Briefly, it involves the use of two dimensional SDS-PAGE to separate the proteins out more than is possible in the one-dimensional system. A denatured sample of cellular proteins is loaded onto a strip gel with an inherent pH gradient and a strong electric field is applied. Under these conditions proteins migrate to the point in the gel whose pH matches their pI (the pH at which the net charge on the protein is zero). This strip gel is then treated with excess SDS so all the proteins are charged in proportion to their mass. The strip is then applied across the top of an SDS-PAGE gel which is run as in the one-dimensional case. Hence the proteome is separated on two axes: charge and molecular weight. In most cases each protein will have its own unique combination of the two parameters and will hence show up as single spot. A comparative study will normally reveal many differences between two cell states or types, some of which will be visible to the naked eye as complete absences on one of the two gels, while others will be quantitative and require computer analysis. The identification of the proteins in spots of interest involves the use of mass spectrometry (reviewed by Gygi and Aebersold, 2000, and Fenyo, 2000). There are two basic methods whose use depends on the organism in question. For organisms whose genome has been sequenced MALDI-TOF (matrix-assisted laser

desorption ionisation time-of-flight) instruments are used. The protein sample is digested by a characterised protease and a mass spectrum of the fragments is produced. This is compared with the theoretical digestion spectra of all the proteins in the genome and an identification is made. When this is not possible, because of the lack of a sequenced genome, the more expensive Q-TOF (quadropole time-of-flight) instruments are used. In this case the proteins are fragmented into multiple random peptides by electrospray and the combination of spectra is analysed to give direct peptide sequence. This latter technique is applicable to *P. fluorescens* until the genome is sequenced.

A further possibility from the proteomic approach is the discovery of protein modifications. In the discussion above it has been implied that the spots of interest will differ in intensity between the gels. However, they may also move position, mainly along the pI axis. Most modifications, such as phosphorylation, methylation and glycosylation, have very small effects on the mass of a protein, but substantial effects on charge. Hence if a protein, such as WspR, was phosphorylated in one genotype but not in another, it might shift on the gel. Not only can such modified proteins be identified, but the nature of the modification can also be discovered. In some cases this is through further mass spectrometry (e.g. Larsen *et al.*, 2001; Schlosser *et al.*, 2001), while in others it depends on databases showing the predicted spot movements for different types of modification (Wilkins *et al.*,1999).

Van Bogelen *et al.* (1999) discuss the particular application of proteomics to microbial organisms. Some examples of successful proteomic analyses in bacteria are the role of CcpA in carbon catabolism in *B.subtilis* (Tobisch *et al.*, 1999), the analysis of the cell cycle in *C. crescentus* (Grunenfelder *et al.*, 2001), and the study of cystic fibrosis strains of *P. aeruginosa* (Hanna *et al.*, 2000). Using similar methods, as well as some of the techniques discussed above, proteomics could prove a very useful tool for analysing the WS phenotype, and specifically the role that WspR plays in it.

One way of using proteomics to understand the role of WspR is to make use of the *wspR* overexpressing genotypes. It can be imagined that there are a certain number of differences that could be seen between 2D gels of the ancestral SM and LSWS, in either the presence of spots or changes in their positions. The ancestor, when overexpressing the locked-on allele *wspR19*, behaves like a WS genotype, however all the factors upstream of WspR will presumably remain ancestral. Conversely, the dominant-negative allele *wspR*9 turns LSWS into a SM. However, all the factors upstream of WspR, where the quenching occurs, will remain as they are in LSWS. Therefore, these two genotypes can be used to distinguish the differences between the ancestor and LSWS that are due to factors upstream of WspR from those that are downstream. This is illustrated in Figure 7.1.

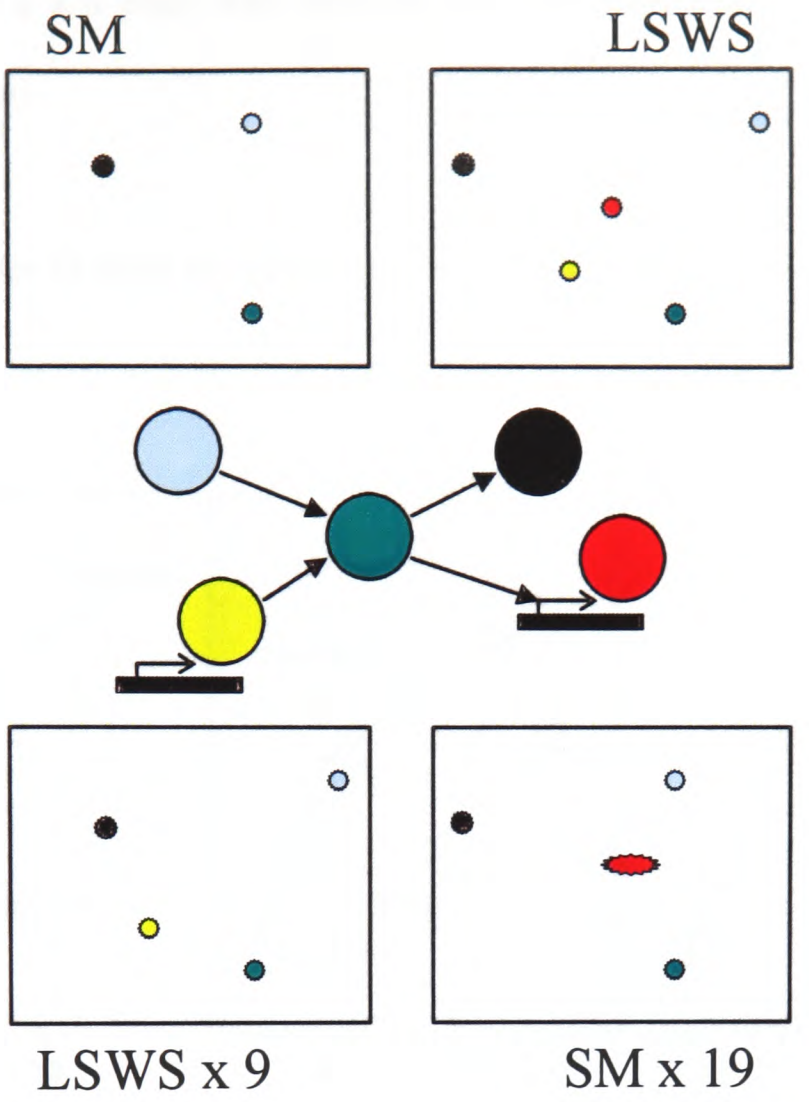


Figure 7.1. Schematic representation of a *wspR* proteomic analysis. This figure shows an idealised version of what might be seen in a *wspR* proteomic analysis. In the centre of the diagram, WspR is represented as a green protein receiving two input signals and causing two output signals in LSWS. In each case one signal is due to protein modification and one to transcription. In the ancestor the red and yellow proteins are not transcribed and the blue and black proteins are not activated by modification. These differences are represented on the two gels above. Both transcriptionally regulated effects are present in LSWS and missing from the ancestor, while the other two proteins move the position of their spots due to an arbitrary modification. The two lower gels show LSWS(pVSP61-*wspR*9) and SM(pVSP61-*wspR19*). The former has the same spots as LSWS on the input side and the same as SM on the output; the latter is the opposite way round. For convenience it has been assumed that the modification to WspR that makes it active does not change the position of its spot.

* * * *

To broaden the study of *wspR* in LSWS, the genotype from which it was isolated, to the study of its role in other WS genotypes, the 26 independently-isolated genotypes described in Chapter 6 were used. In some of the assays performed in this chapter all of the genotypes were used. However, for other assays a subset was required for practical reasons, but it was still hoped to represent the entire range. Six genotypes were selected on the basis of their

ability to spread across a KB plate with different agar concentrations (A. Spiers and P.B. Rainey, unpublished data).

The spreading phenotypes of these six genotypes, WSe, WSi, WSn, WSs, WSv, and WSx is shown in Figure 7.2.

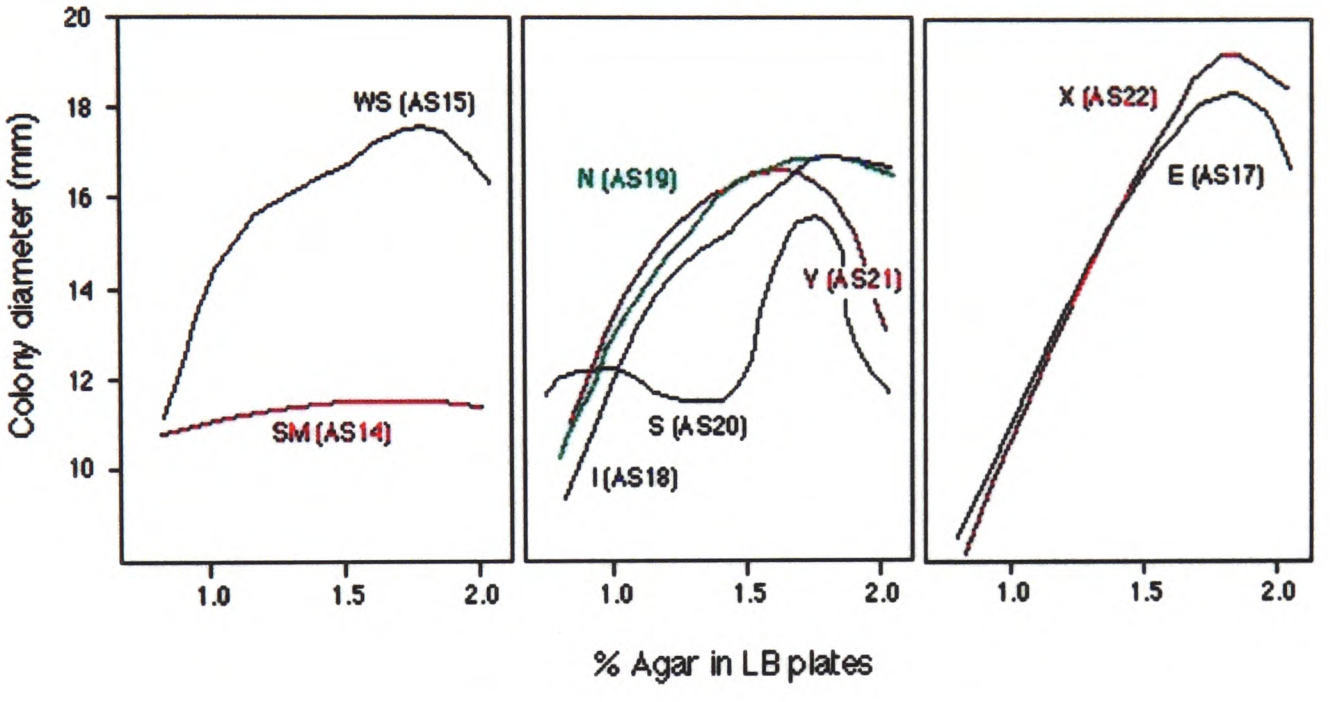


Figure 7.2. Spreading ability of WS genotypes. The relationship between colony diameter on KB plates and the agar concentration is shown for SM and LSWS (first graph), WSi, WSn, WSs and WSv (second graph), and WSe and WSx (third graph). WSi, WSn and WSv are most similar to LSWS, though they do differ. WSe and WSx belong to a separate class and WSs is unique. None of them are similar to the ancestor. Figure obtained from A. Spiers (personal communication).

The *wssE-lacZ* fusion, used in SM and LSWS in Chapter 3 (genotypes AS14 and AS15 respectively), was introduced into these six genotypes so that similar assays of *wss* transcription could be made. This was one assay which was selected for use in studying these genotypes, and the effects of *wspR* on them, in this chapter. Two new assays, one of attachment (the initial stage in biofilm formation) and a quantitative Congo Red assay for cellulose, were also performed. In order to make connections between these phenotypic assays and fitness, as was done in Chapter 5, fitness assays were also performed. In this way an assessment of the WS phenotype and its variation could be made.

7.2 Results

7.2.1. The relationship between motility and the LSWS phenotype

To assess the role of WspR in motility, five genotypes were subjected to a motility assay, as described in Chapter 2. These were the non-motile strain, NM, as a negative control, the ancestral SM and the LSWS, and two mutants, the *wssB* mutant WS13 and the *wspR* mutant WS4. It was hoped that any differences between the ancestor and LSWS could thus be attributed to the action of either *wspR* or the *wss* operon. The results of the assay are displayed in Table 7.1.

Strain	Average Swarm Diameter (mm)	SE
NM	0.0	0.00
SM	15.6	1.11
LSWS	1.1	0.11
WS4	15.8	2.47
WS13	1.2	0.14

Table 7.1. Motility of WS mutants. The average swarm diameter from 5 replicates grown in 0.35% agar KB plates for 24 hours is shown with standard errors (SE). WS4 and WS13 are *wspR* and *wssB* mutants respectively.

The results show that the LSWS has a severely reduced motility compared to the ancestor, though it is not completely non-motile like the control. The low-motility phenotype of LSWS is rescued (to SM levels) by the WS4 mutation, but not by the WS13 mutation. This shows that it is the over-activity of *wspR* in LSWS that hinders motility, not that of the *wss* operon.

7.2.2. 2D gel analysis of the LSWS phenotype

As explained above, the proteomic approach has the potential to identify the global changes associated with the SM to WS transition. Also, by analysis of certain *wspR* overexpressing genotypes, it could potentially identify proteins that are upstream and downstream of WspR. In an initial attempt to address the first question SM and LSWS samples were run on an 11cm pH3-11 linear immobiline strip, and subsequently onto a second dimension gel prepared in

the laboratory, as described in Chapter 2. Visual inspection showed no differences in spot patterns between the two samples. As most of the proteins were found in the centre of the gels it was decided to use a non-linear strip, in which the pH5-8 region is expanded so that a finer level of detail can be observed. This time four samples were prepared: SM and LSWS, the locked-on genotype SM(pVSP61-*wspR19*), and the dominant-negative genotype LSWS(pVSP61-*wspR*9) (for reasons explained above). On this occasion 7cm non-linear pH3-11 strips were used and the second dimension was run on a pre-cast minigel. The results are displayed in Figure 7.3. and the obvious differences in spot pattern are marked. Due to the lack of replication and the less than optimal quality of the gels no computer analysis was attempted. Inspecting by eye it is clear that the phenotypically SM genotypes have two spots in region A, compared with the phenotypically WS genotypes' one. Region B is less clear, though the pattern in the two chromosomally ancestral genotypes seem similar, and likewise with those that contain the LSWS chromosome. As yet, these spots have not been subjected to Q-TOF analysis, though the current sequencing of the *P. fluorescens* SBW25 genome (P.B. Rainey, personal communication) may allow MALDI-TOF analysis to be performed instead.

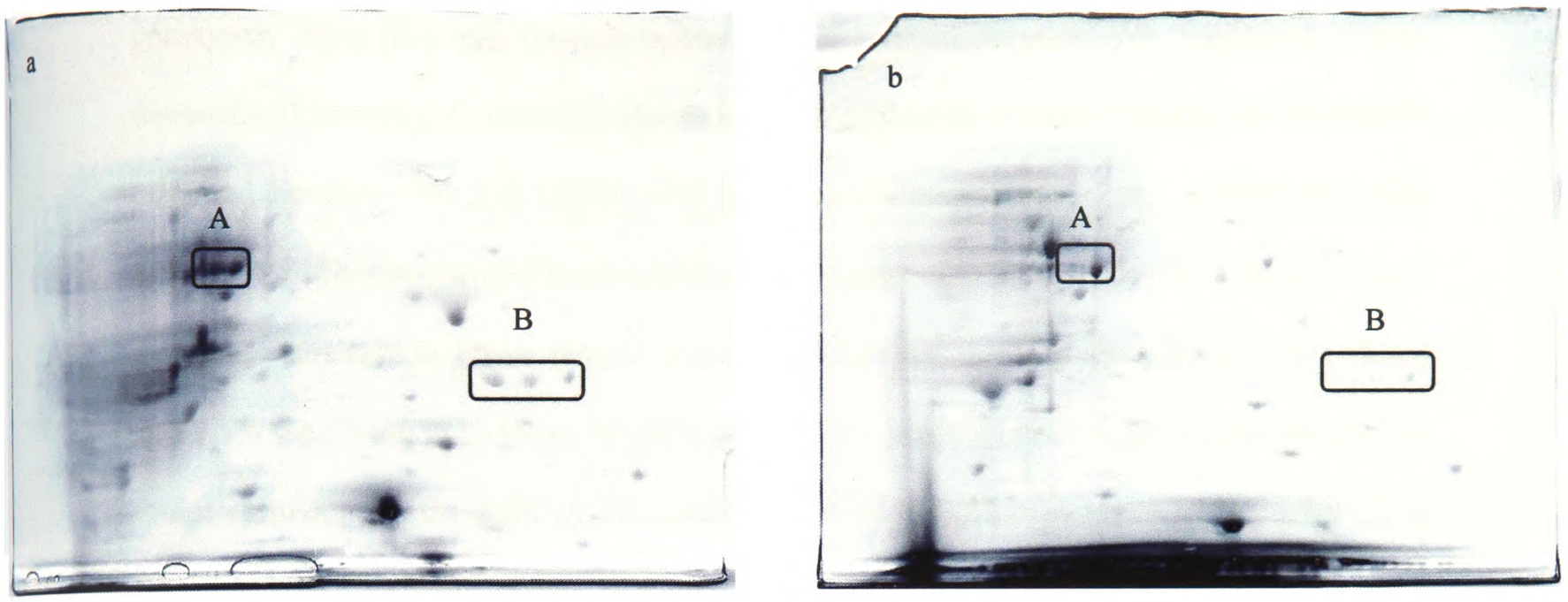


Figure 7.3. 2D-PAGE of whole cell extract from the following genotypes: a) SM; b) LSWS; c) SM(pVSP61-*wspR19*); d) LSWS(pVSP61-*wspR9*). High molecular weight proteins are at the top of the gels, and pH3 is at the left hand side. Region A is marked on all four gels showing samples a and d with 2 heavy spots while b and c have only one. This is the clearest difference between the gels. Also marked is Region B where each sample has a different weighting of the 3 spots.

7.2.3. 1D gel analysis of independently isolated WS genotypes

The differences shown on the 2D gels above were all between ancestral and LSWS genotypes. As a first step towards extending the analysis to other WS genotypes, a one-dimensional protein gel of the 26 independent WS genotypes was run. Though the differences observed between SM and LSWS were all second-dimensional, it was possible that first dimension differences would be seen with a wider range of WS genotypes. In order to achieve maximal visualisation across the gel, different concentrations of polyacrylamide were tested and 7.5% was found to be ideal. No differences were apparent between genotypes, though this is not surprising in the light of the results above. If the 2D analysis of SM and LSWS is typical, and all differences are in the pI dimension rather than that of molecular weight (implying that the same proteins were present but in different states.), then none would be seen on a one-dimensional gel.

7.2.4. The introduction of *wspR* alleles into 6 WS genotypes

Six WS genotypes were selected as being representative of the range of phenotypic variation. The reasons behind this are discussed above. The pVSP61vectors containing *wspR9*, *wspR12* and *wspR19*, along with pVSP61 itself as a plasmid control, were conjugated into the six genotypes, WSe, WSi, WSn, WSs, WSv, and WSx. As a preliminary step to performing other assays, the colony morphologies on KB agar were recorded. As expected *wspR12*, *wspR19* and the control all resulted in WS colony morphology, though there was considerable variation within this class. The effect of the dominant-negative allele, *wspR9*, however, was less intuitive. While it caused its expected phenotype (SM colony morphology) in four of the genotypes, it failed to produce a dominant negative effect in WSs and WSv. To test whether this effect was specific to this allele, or was general property of these two genotypes, four more dominant-negative alleles were conjugated into the six genotypes. These alleles were *wspR1*, *wspR5*, *wspR13*, and the liberated N terminus, *wspR*NtermB. These all produced a dominant-negative effect in WSe, WSi, and WSn, and all failed to produce such an effect in

WSs and WSv. The more complicated case was WSx, in which *wspR*1 and *wspR*13 were dominant negative, *wspR*5 gave a variant smooth morphology, and *wspR*NtermB failed to reverse the phenotype. These results will be discussed further below, but for the subsequent experiments in this chapter *wspR*9 will be used to represent the dominant negative alleles.

7.2.5. *Wss* operon transcription

In Chapter 3 the level of *wss* transcription was found to vary little between SM and LSWS, and also on the addition of *wspR* alleles (with one exception). However, this does not mean that other WS genotypes do not vary in their levels of *wss* transcription. Hence a β-galactosidase assay was performed on the *wssE::lacZ* versions of the six genotypes and also on the same genotypes overexpressing the *wspR* alleles. No large effect was expected with the alleles because they did not greatly affect LSWS, and also because WspR does not act at a transcriptional level. Unfortunately the comparative data for SM and LSWS genotypes was not available and it is not possible to compare these results with the assay in Chapter 3 because the variation between experiments is likely to be large. The data is displayed in Figure 7.4. Two-way ANOVA shows that there is significant variation in *wss* transcription due to both genotype ($P<0.0001$) and *wspR* allele ($P<0.0001$), as well as a significant interaction between the two ($P<0.0001$). There is variation between the plasmid control samples representing different native levels of *wss* transcription in the respective genotypes. There is also a difference in the response to *wspR* alleles with WSe, WSi and WSn having similar patterns; WSx being slightly different and WSs and WSv both being unique. WSs is the only genotype in which any *wspR* allele increases the level of *wss* transcription above that of the control.

Figure 7.4. Beta-galactosidase activity of six WS genotypes a with *lacZ* fusion to *wssE* overexpressing *wspR* alleles.

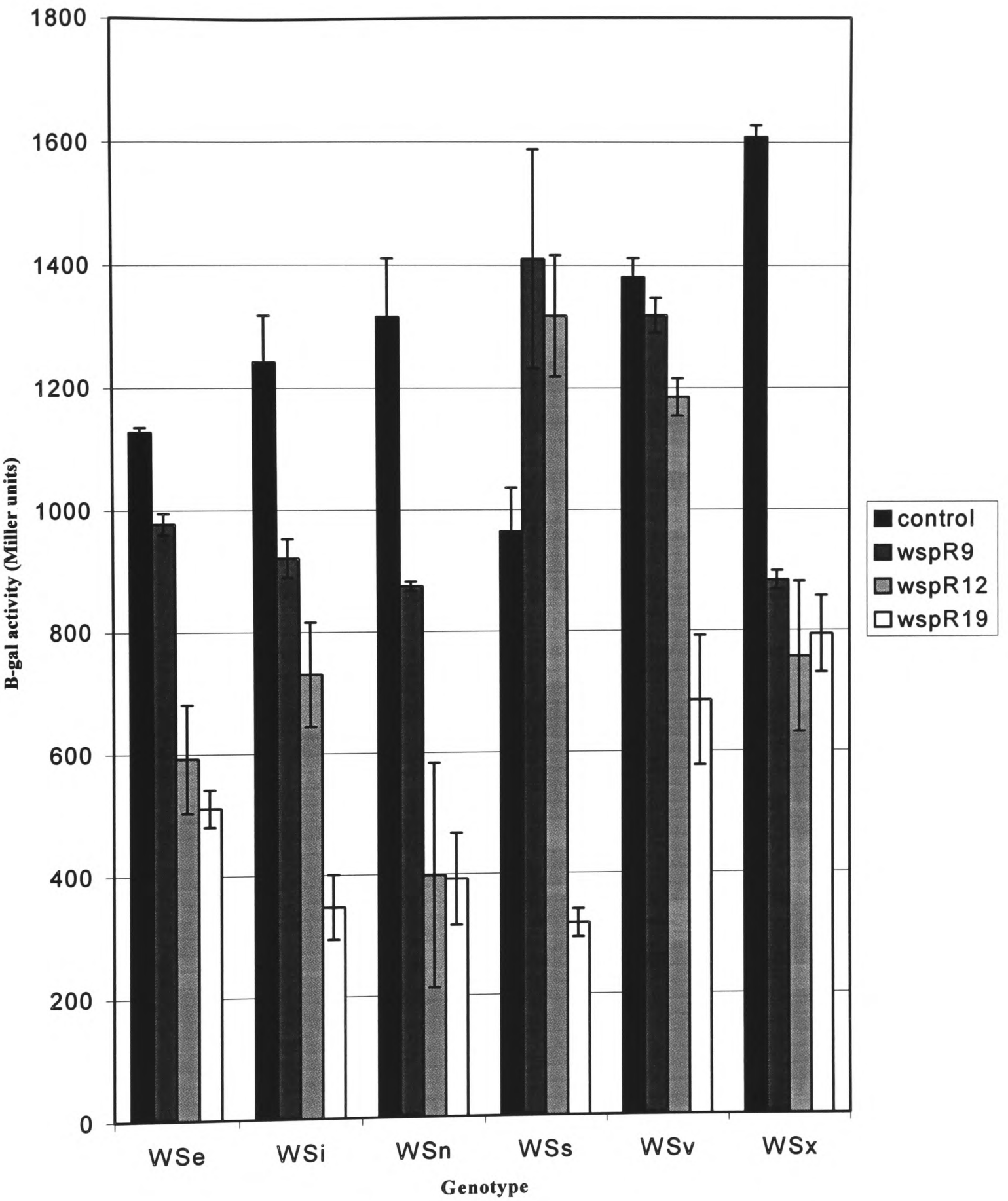


Figure 7.4. β-galactosidase activity of six WS genotypes overexpressing three *wspR* alleles and a plasmid control. The activity in Miller units is shown for each genotype-allele combination (using genotypes with a *lacZ* fusion to *wssE*). The assays were performed on LB broth cultures that had been grown overnight. Each sample was assayed in triplicate and the standard error is shown.

7.2.6. Attachment assays

Attachment is the initial step in biofilm formation and, as discussed in Chapter 1, often requires the action of pili and flagella. Therefore an attachment assay was performed on the six WS genotypes with each of the *wspR* alleles. It was expected that all the genotypes would attach more than the ancestor, and it may be that a locked-on copy of WspR, with its potential flagellar interaction, can increase this further. The effects of the dominant-negative allele (*wspR9*) are also interesting, as it can quench WspR activity, but would not be able to quench any increased attachment that has occurred independently of WspR. The results for the six genotypes and four alleles, along with SM and LSWS controls for reference, are shown in Figure 7.5. Two-way ANOVA shows that there is significant variation due to both genotype (P=0.0475) and *wspR* allele (P<0.0001), as well as a significant interaction between them (P<0.0001). As expected, LSWS attaches significantly more than SM, as do most of the other WS genotypes. Four of the genotypes respond strongly to *wspR*12 and *wspR*19 but WSs and WSv show no significant response to any *wspR* allele. These are the same genotypes that do not show a dominant-negative response.

Figure 7.5. OD570 of attachment samples for six WS genotypes overexpressing four *wspR* alleles with SM and LSWS controls.

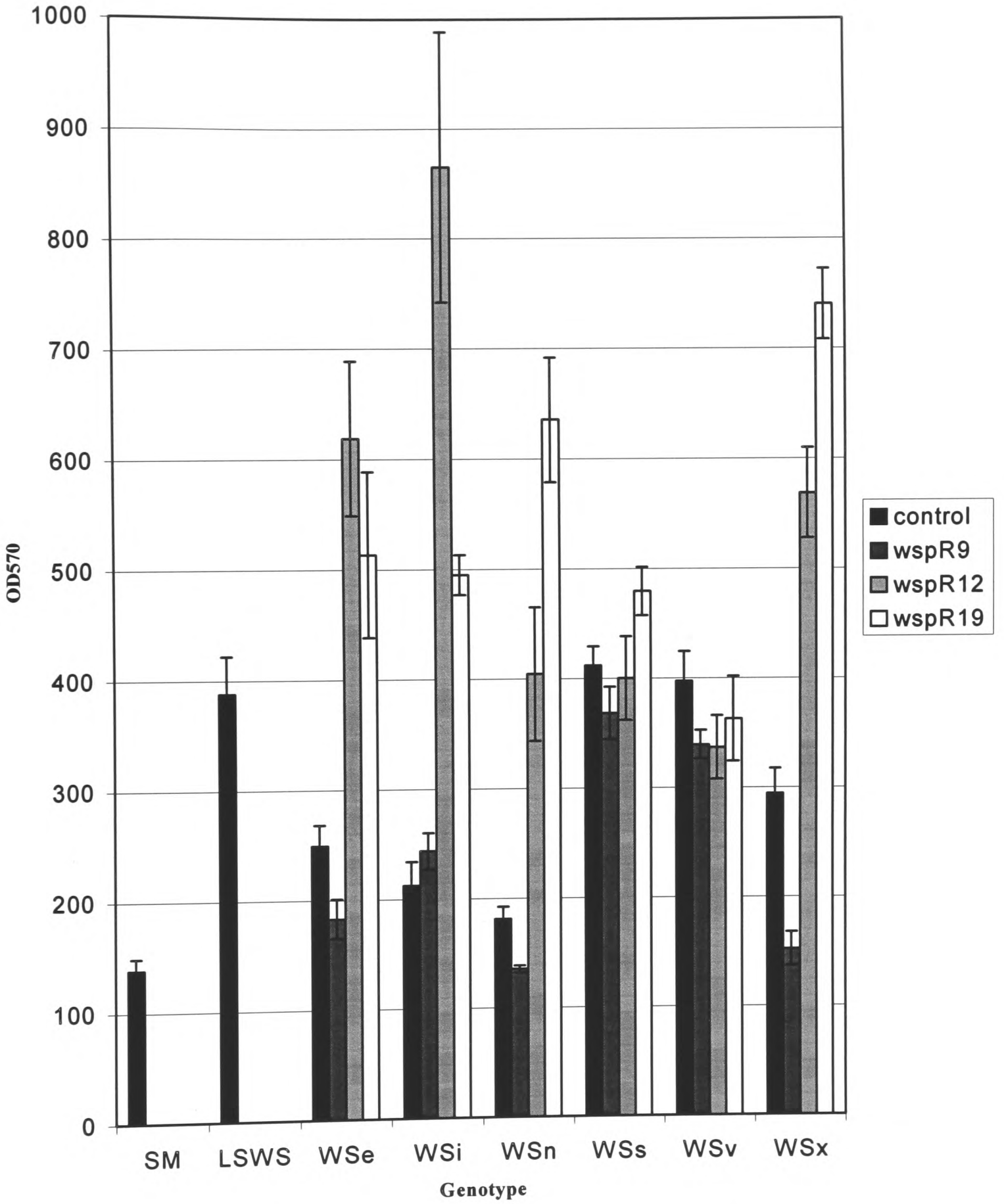


Figure 7.5. Attachment of six WS genotypes overexpressing three *wspR* alleles and a plasmid control. The attachment is measured as the OD_{570} of the crystal violet solution after removal from the microcosms. This is shown for each genotype-allele combination. The assays were performed on KB microcosms that had been grown overnight. Each sample was assayed in triplicate and the standard error is shown. SM and LSWS genotypes, with the plasmid control are also shown.

7.2.7. Quantitative cellulose assays

While there is a qualitative difference in cellulose production between the ancestor and all WS genotypes, there may also be a quantitative difference between these genotypes themselves. This might reflect the different mutational routes by which the phenotype has been reached. It is likely that the dominant *wspR* alleles will increase cellulose production, so long as the maximum level of biosynthesis possible is greater than that already occurring in the genotype in question. On the other hand, the dominant-negative alleles should only decrease cellulose in the genotypes for which they affect colony morphology. The results of the assay are shown in Figure 7.6. Two-way ANOVA shows significant variation due to *wspR* allele (P=0.0311), but not due to genotype (P=0.239). However, a one-way ANOVA, performed on just the plasmid controls (hence representing the native levels of cellulose production in the genotypes without the effects of *wspR* alleles) showed a significant variation due to genotype (P=0.00257). The data for *wspR19* in all genotypes and for WSi + *wspR12* was too inconsistent and is therefore not shown, though it was included in the analysis of variance. There is an obvious and significant difference between LSWS and the other three controls, as expected, and there are noticeable effects of some *wspR* alleles which will be discussed below.

Figure 7.6. Quantitiative Congo Red assays of 6 WS genotypes overexpressing *wspR* alleles and controls.

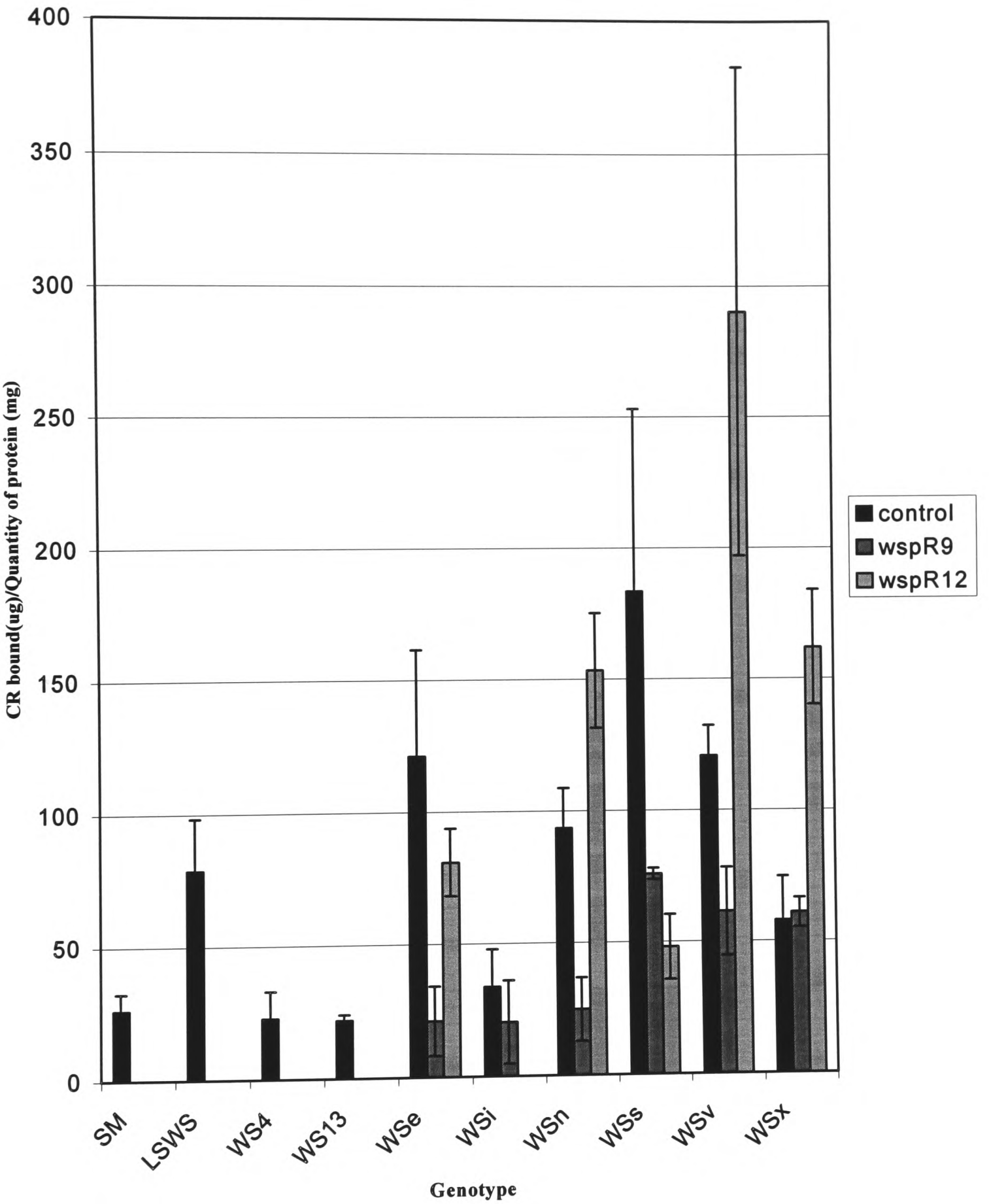


Figure 7.6. Congo Red binding of six WS genotypes overexpressing three *wspR* alleles and a plasmid control. The quantity of Congo Red bound (representing the amount of cellulose) per mg of cellular protein for each genotype-allele combination is shown. The assays were performed on LB broth cultures that had been grown overnight. Each sample was assayed in triplicate and the standard error is shown. SM, LSWS, WS4 and WS13 genotypes are also shown. No data is shown for WSi(pVSP61-*wspR12*).

7.2.8. Fitness

As discussed in Chapter 5, the most important feature of the WS phenotype is its enhanced fitness in comparison with the ancestor. All of the phenotypic assays performed above assess potential components of fitness, but a fitness assay itself is required to be able to understand their importance. Therefore a fitness assay on the 6 WS genotypes was performed for 24 hours in a static microcosm in competition with the ancestor. This is the most basic of the possible competitions, and is identical to that performed in Sections 5.2.1 and 5.2.2 for LSWS and the ancestor. The results of the assay are displayed in Figure 7.6. Two-way ANOVA shows that there is significant variation due to both genotype (P=0.0287) and allele (P=0.000492), as well as a significant interaction between them (P=0.00633). It is notable that, contrary to expectation, few of the genotypes have an advantage over the ancestor (LSWS and WSi are the exceptions) and that each genotype appears different in its response to *wspR* alleles. WSs and WSv, the genotypes that showed no dominant-negative response and no response to *wspR* in the preceding assays, show no significant fitness response to *wspR* either.

Figure 7.7. Fitness of SM, LSWS, and 6 WS genotypes overexpressing *wspR* alleles competed against SM*panB* for 24 hours in a spatially structured environment

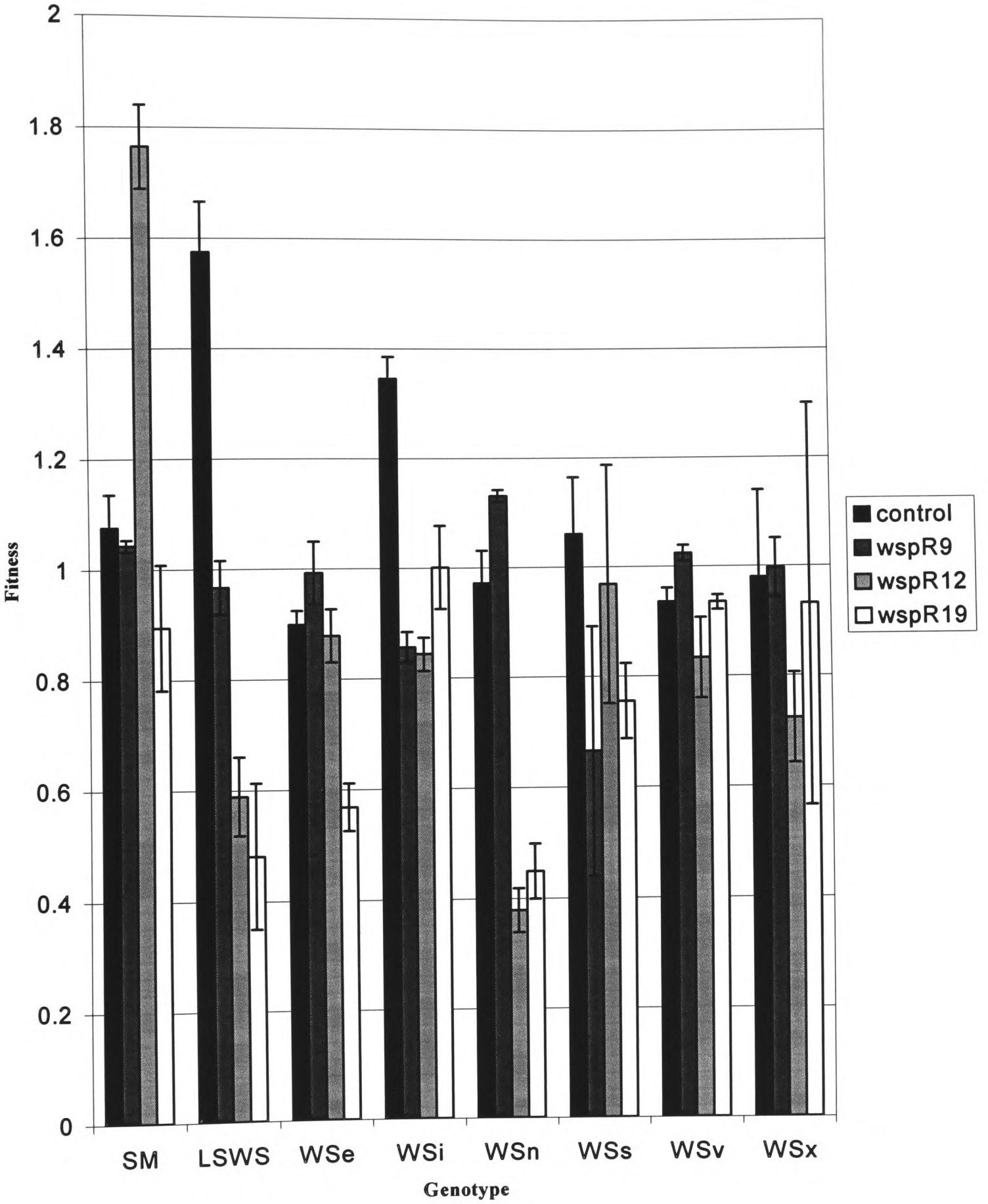


Figure 7.7. The fitness effects of three *wspR* alleles and a plasmid control on six WS genotypes and LSWS and SM controls in a structured environment. The fitness of each genotype-allele combination measured in competition with SM*panB* over 24 hours in a static KB+pan environment. In each case three replicates were performed and the error bars represent the standard error.

7.3. Discussion

7.3.1. WspR and motility

The motility assay shows very clearly that the ancestor is motile and that LSWS has this ability substantially impaired. It is not, however, quite as impaired as the non-motile genotype. While this latter genotype has not been characterised it is likely that it is deficient in some essential motility-related structure such as the flagella. The fact that LSWS is slightly less impaired may be because it is deficient in some regulatory function further upstream in the pathway. The assay also studied the WS13 and WS4 genotypes in an attempt to uncover the reason for this deficiency. It may be that the overuse or misuse of the flagella in biofilm formation results in an inability to swim (Deziel *et al.*, 2001). Alternatively, it could be that the overproduction of cellulose itself impairs motility in some unknown way, possibly by causing the cells to adhere to each other. This second possibility is ruled out by the fact that the WS13 mutation does not rescue motility. This mutation is in *wssB* and is polar so cellulose is not produced. The WS4 mutation, on the other hand, does rescue motility implying that WspR, or something downstream of it other than the *wss* operon, is responsible for the loss of motility. This implies very strongly that WspR interacts with other factors and is therefore a global regulator, rather than merely a regulator of cellulose synthesis. This seems logical in light of the fact that PleD is a regulator of flagellar degradation and hence has a role in motility (Aldridge and Jenal, 1999). These results do not allow a distinction between the possibility that WspR regulates motility in a negative way, hence its overactivity diminishes motility, and the possibility of it affecting motility-related structures in a positive way hence causing diminished motility by their inappropriate overactivity.

7.3.2. Proteomic analysis

The proteomic analysis carried out in this chapter was very preliminary in character and therefore limited in its scope. All the gels have a relatively low number of distinguishable spots compared with the thousand or so that might be expected. This may be due in part to an inefficient sample preparation in which the recovery rate of protein was as low as 10%. A more efficient way would be the use of sonication to lyse the cells rather than the low-dosage SDS treatment that was used here. It is hoped that the method used was merely inefficient and not selective in which proteins were obtained. The quality of the gels is not ideal with smearing near the acidic end and there was no replication of samples. Ideally there would be at least three replicates of each gel, especially if software was going to be used to identify quantitative as well as qualitative differences. Most significantly the analysis was not taken to the stage of spot identification by mass spectrometry, though this possibility remains open for the spots of interest that were reported.

Despite all these caveats the analysis did achieve its major aim which was to show differences between SM and LSWS genotypes at the proteomic level. There are undoubtedly more that cannot be identified because the gels are not sensitive enough, or because the sample preparation was selective, but, crucially, some were identified. The most immediately obvious difference is that marked as region A in Figure 7.3. In this region there are two spots of equal size but slightly different pI. Both these spots are present in SM and LSWS(pVSP61-*wspR9*) (the two phenotypically SM genotypes) whereas the phenotypically WS genotypes possess only one of them. This pattern of spot presence between the genotypes corresponds to the factors downstream of WspR discussed in Figure 7.1. If they are the same protein that has been modified then they would correspond to the spot that has moved, though the case is more complicated here because both forms exist in the SM genotypes. Alternatively the mass identity between the two spots could be coincidental, and the one that is present in SM and absent in WS is the only one of interest. In this case the protein would correspond to one that was transcribed in one genotype and not the other. Unlike Figure 7.1, however, it would be present in SM and absent in WS, implying that WspR is affecting its transcription negatively.

The only way to resolve the difference between these possibilities would be identification by mass spectrometry. Some speculation is possible, however. The spots are relatively large which would imply a structural rather than regulatory protein. A suitable candidate might be a flagellar component that exists in two states, one of which is inactive or destined for degradation. This form is not present in the constitutively active WS genotypes.

The differences between the gels in the region marked B are a lot less clear, and each gel appears to have a slightly different pattern. Tentatively, however, it seems that the genotypically SM genotypes have a larger amount of the two acidic versions, whereas the LSWS genotypes have more of the most alkaline one. If this were indeed the case the spots would correspond with factors upstream of WspR, as shown in Figure 7.1. Again the interpretation depends on whether the spots are modified versions of the same protein or different proteins altogether. Assuming that at least two of the spots are the same protein they could represent the histidine kinase or the MCP, whose level of activity differs between SM and WS due to modification (phosphorylation in the former case and methylation in the latter). Regulatory proteins are usually found at much lower levels than structural proteins so they are often not visible on SDS gels; this might be a particular problem if the sample preparation is inefficient, as it is here. However there is evidence (Bantinaki, 2001) that the *wsp* operon is transcribed (though not necessarily translated) at very high levels in both SM and LSWS, which could explain this result.

The differences that were discovered by 2D analysis would probably not have been apparent on a one-dimensional gel as they were differences on the pI axis, and indeed they were not even visible on a 2D gel with a linear pI axis. If this is generally true of the differences involved in the WS phenotype it is not surprising that the 1D analysis of different WS genotypes showed no differences. A larger 2D analysis of these genotypes is now underway (P.B. Rainey, personal communication) that hopes to uncover such differences. These will be identified by Q-TOF mass spectrometry in the first instance, though once the genome project

is complete, MALDI-TOF will become an option. If the speculation on the nature of the spots identified above is correct, then some differences should be apparent, especially if a quantitative analysis is undertaken. The differences in upstream regulatory elements seen between SM and LSWS will also be seen between different WS genotypes if they have used different mechanisms to reach the WS phenotype. On the downstream side, it is unlikely that there will be qualitative differences in the cellulose synthase genes, but there may be quantitative ones. Differences in flagellar and other motility-related proteins will be seen if they are not essential for the WS phenotype and are merely side-effects of increased WspR activity. This depends on the relationship between fitness, attachment, and cellulose biosynthesis. Hence this approach is likely to provide substantial information about how the different aspects of the WS phenotype relate to each other and which are essential and which are incidental.

7.3.3. The effect of *wspR* alleles on WS genotypes

The action of the dominant-negative alleles of *wspR* has been discussed at length in Chapter 4. It was concluded that when overexpressed they were acting as phosphate sinks quenching the enhanced signal coming through the *wsp* operon. This effect would therefore be expected in any WS genotype where *wsp* activity had been enhanced in order to achieve the WS phenotype. However, any genotype that has a mutation that causes the WS phenotype without increased *wsp* activity would not be quenchable in this way. The results of introducing *wspR9* into the six independent WS genotypes implies that WSs and WSv fall into this latter category, while the other genotypes involve *wsp*-related mutations. This was confirmed by the addition of other dominant-negative alleles, with the exception of WSx. This genotype behaved in an allele-dependent way with one allele failing to quench at all. This allele was the liberated N terminus which might be expected to be the least efficient phosphate sink for the structural reasons discussed in Chapter 4. This is especially likely in light of the fact that the other liberated N terminus (*wspR*NtermA), which failed to quench LSWS, is only 5 amino

acids shorter. In this case it may be that WSx has enhanced *wsp* activity causing the WS phenotype, but that it is sufficiently enhanced that a weak phosphate sink is unable to quench it.

This suggestion of no *wsp* involvement in WSs and WSv, if true, might be expected to have implications for other aspects of the phenotypes of all the WS genotypes, and also the effect that *wspR* alleles have on them. The attachment assay shows the clearest difference between WSs and WSv, and the other genotypes. There is a range of attachment abilities between the different WS genotypes (as seen by looking at the plasmid controls). It is likely that there is a trade-off between enhanced attachment and the resultant loss of motility, and that different WS genotypes have reached a different compromise. The four genotypes WSe, WSi, WSn and WSx respond to *wspR* much as would be expected, with increased attachment due to *wspR12* and *wspR19*, and in most cases a decrease with *wspR9*. The former is probably through the effect of WspR on motility while the latter is due to quenching. The other two genotypes do not respond to any *wspR* alleles, positive or negative. It might have been expected that WspR19, at least, would have an effect on these two genotypes as it is signal-independent. However, if they have mutations affecting attachment more directly they may not respond to WspR regulation at all, positive or negative; this would also explain them having the highest native levels of attachment. Another possibility is that attachment and cellulose production are linked by WspR activity in the four genotypes that do respond to the alleles, and thus the two factors impose fitness constraints on each other. It is possible that LSWS and these four genotypes attach sub-optimally because cellulose production is very high and there is a trade-off. On the other hand WSs and WSv may have separate mutations affecting each. Presumably one mutation occurred to give the initial fitness advantage followed by the other to enhance it further. Furthermore, the fact that WSx has the highest attachment of the other genotypes fits with the suggestion above that it has the highest level of *wsp* activity.

The quantitative Congo Red assay is unfortunately less interesting due to data for some samples in which the standard error was almost as large as the mean. The data that remains is, however, of interest, especially as there is a significant effect of *wspR* allele. The LSWS control produces three times as much cellulose as the ancestor, and both the WS4 and WS13 mutants revert to the ancestral level as expected. Beyond that it is difficult to speculate about genotype effects as the two-way ANOVA showed them as not being significant. Genotypes WSe, WSi and WSn show the expected result when *wspR9* was overexpressed, reducing the cellulose production to ancestral levels. The addition of *wspR9* to WSs and WSv also decreases cellulose production, albeit only to the level of LSWS. This seems odd if they are not susceptible to quenching; also WSx does not respond to *wspR9* at all. This could be because in WSs and WSv the expression levels of the *wsp* operon, rather than the activity of individual components, have been increased. This alternative explanation would make quenching to SM levels of CLP impossible because the alleles are no longer overexpressed in relative terms, though quenching to LSWS levels might be possible. This explanation is very different from the one offered above that suggests no *wsp* involvement in the two genotypes WSs and WSv. In this case these genotypes would have enhanced cellulose production through increased *wsp* operon expression which can be partially quenched by WspR9, but they have increased attachment by a separate mechanism which cannot be quenched. This is, however, highly speculative especially given the large errors in the Congo Red assay. This assay does give the expected result with respect to *wspR12*, which causes an increase in cellulose production in all genotypes.

In the fitness assays the two genotypes, WSs and WSv, also show little response to *wspR* alleles. This would be consistent with overexpression of the *wsp* operon drowning out the effect of the overexpressed *wspR* alleles. It is interesting that, aside from LSWS, the only genotype with a significant fitness advantage over the ancestor is WSi; the other genotypes must rely on a frequency dependent effect in niche colonisation. As they have fitnesses near to unity it is no surprise that *wspR9* does not appear to reduce them, affecting only WSi and

LSWS. The dominant alleles, *wspR12* and *wspR19* have the expected negative effects on WSe, WSi and WSn, similar to the effects on LSWS. The data for WSx is too unclear to analyse and it could be behaving like WSs and WSv, or the other genotypes. Each genotype appears to have its own unique pattern of fitness reactions to *wspR* alleles.

The analysis of *wss* operon expression levels of SM and LSWS genotypes in response to *wspR* alleles, presented in Chapter 3, showed few changes, especially in LSWS. It is therefore surprising to see such significant changes in this assay. There are some differences between the genotypes containing the control alleles, corresponding to natural variation in *wss* expression. It is the introduction of *wspR* alleles, however, that make the more pronounced differences. A small decrease in expression is expected with *wspR9* when compared with the data from Chapter 3. Such a decrease is seen here (though it is large in some cases) and it was suggested in Chapter 3 that it might be due to positive feedback of the *wss* operon on its own transcription. This decrease is not seen in WSs and WSv, as would be expected if quenching does not occur. The effects of the dominant alleles, which mostly decrease transcription dramatically, are harder to explain. There was no evidence in Chapter 3 for negative feedback with these alleles. These data are very difficult to interpret without the proper SM and LSWS controls for comparison.

In this last assay WSs and WSv again have different patterns of responses to *wspR* alleles from the other genotypes. When taken together these assays strongly imply that these two genotypes are qualitatively different from the others. The explanation for this is not clear, though some highly speculative suggestions have been offered above, regarding either a lack of *wsp* involvement, or a change in *wsp* expression levels rather than activity. It is interesting to note that there seems little to distinguish these two genotypes from the others in the spreading assay that was the basis of their selection for study (Figure 7.2). This may be because the assay was purely artefactual, and it was serendipity alone that a range of phenotypes was chosen. As already mentioned, the analysis of these genotypes and the effect

of WspR on them, gives a unique profile of each genotype, and leads to much speculation on the mechanistic explanations for the differences between them. The predictions that come from this will be tested in the mutational and proteomic studies that have already begun (P.B. Rainey, personal communication).

Chapter 8

General Conclusion

8.1. Summary of major findings

The mutagenesis of LSWS carried out by Kahn (1998), and specifically the WS4 mutant, demonstrated the necessity of *wspR* for the phenotype. Work carried out in this study and by others (Spiers *et al.*, 2002; Bantinaki, 2001; J. Bohannon & P.B. Rainey, unpublished) go further towards demonstrating the mechanistic basis for this necessity. However, this study has also demonstrated the extent to which WspR is sufficient for the various aspects of the WS phenotype. The overexpression of wild-type *wspR* in the ancestral SM genotype demonstrates that, under certain circumstances, a WS phenotype can be created by this alone. If it is assumed, as seems likely, that overexpression causes an increase in the level of WspR activity, the crucial importance of this protein has been demonstrated, as its activity level alone determines the phenotype. This is probably the case in other WS genotypes as well. This is in contrast to the effect of overexpressing the *wss* operon which is not sufficient for the WS phenotype. The caveat associated with this sufficiency is the requirements for salt for the WS colony morphology (though not so much for the fitness effects). This observation has been put to use in formulating a model of WspR activity, the predictions of which were tested in Chapter 4. It was also found that the sufficiency of WspR could be 'improved' with the mutations to the N terminus. These were found to be signal-independent, thus removing the need for salt. As well as giving useful insights into the functioning of the protein they are useful tools for future research and are the source of interesting evolutionary data in Chapter 5.

The specific mutations introduced into *wspR* (described in Chapter 4) elucidated several aspects of its mechanism. The role of phosphorylation is demonstrated by the requirement of aspartate-67 for activity. The importance of WspR as a rate-limiting component in the

signalling pathway was shown by the domain liberation experiments which also demonstrated the modular nature of the protein. This modularity was exploited in the creation of WspR-PleD hybrids. These confirmed that the similarity between the sequences of the two genes corresponds to a similarity in mechanism of action, although the need for hybrids rather than wild-type proteins, and the unidirectional nature of the complementarity, demonstrate that there are important differences between the two systems. It remains unclear how this relates to the possible role of cyclic-di-GMP in both systems.

As this study aimed to uncover molecular information about an evolutionary event it is appropriate that information should also travel in the opposite direction. In Chapter 5 the evolutionary consequences of the molecular aspects of *wspR* were studied. The rationale behind this is to use artificial manipulation to learn about the nature of selection in the microcosm. This approach, with the use of manipulations in a candidate gene, is similar to that of Feder (1999) in *D. melanogaster*. The results demonstrate that overexpression of *wspR* alleles can increase the fitness of the ancestor, thus making them potential targets of selection. It also confirms that trade-offs exist between fitness in different environments, but that phenotypic plasticity can circumvent this: the plastic genotype SM(pVSP61-*wspR12*) seems able to switch between biofilm and planktonic growth to achieve maximal fitness. The existence of this, artificially-created, fit genotype illustrates the existence of constraints acting during microcosm evolution. Such a genotype would be expected to have arisen amongst the collection of twenty six WS genotypes if a constraint did not exist. However, the fact that the genotype only occurs when created by artificial manipulation implies that there is a constraint. The use of different environments and conditions (e.g. starting frequencies, time-scale) show that fitness is a complex phenomena that depends on specific parameters as some genotypes are unfit over short periods or when common, but have long-term or frequency-dependent advantages. Possibly the most important outcome of these results is their sometimes counter-intuitive nature. On the basis of the phenotypes described in the Chapters 3 and 4 certain fitness effects were expected, such as an enhancement of that of the ancestral

SM by the dominant alleles. This was the case with one allele (*wspR14*) but not with the other (*wspR19*), though here again the environmental conditions were crucial. This is an important conclusion for studies that look at candidate mutations for adaptive evolution (Feder and Krebs, 1998; Tatar, 1999). Without making direct assessments of fitness, the distinction between adaptive and maladaptive mutations is difficult to assess; In prokaryotes this is often a simpler task than in multicellular organisms as fitness can be assessed directly by the ratio of the Malthusian parameters (Lenski *et al.*, 1991).

The rationale behind the focussed study of *wspR* was that it is a candidate gene for evolutionary change. This was confirmed by the effects on phenotype and fitness that arose from alterations in *wspR* sequence and expression described in Chapters 3, 4, and 5. In Chapter 6 *wspR* was sequenced from independent WS genotypes to discover whether this potential evolvability corresponded with actual sequence variation. The answer was that it did not. This may say something about the genetics of adaptation as a predicted mutation does not occur despite a known selective advantage. However, none of the *wspR* mutant alleles were introduced into the chromosome in single-copy, though the differences in the severity of the different alleles in multi-copy imply that some of them would produce the desired fitness increase if introduced into the chromosome. Further evidence of a mechanistic constraint was given in the analysis of the plastic genotype which does not evolve unless the level of *wspR* transcription is uncoupled from that of the rest of the *wsp* operon.

The tests of the effects of *wspR* alleles on different WS genotypes (Chapter 7) highlighted two of the six genotypes as being markedly different in their response. Unlike the other genotypes these did not show major effects of *wspR* overexpression in the assays performed. Though the precise reason for this remains unclear it clearly demonstrates a significantly different mutational route to the WS phenotype from the one found in the LSWS genotype. This is not a surprising result, but what is unclear is whether the different routes are equivalent in terms of fitness and taken at random, depending merely on which mutation arises first, or whether

different WS genotypes co-exist in the same microcosm population as they are differentiated by frequency-dependent selection. Rainey and Travisano (1998) found a range of WS phenotypes within each microcosm, making the latter possibility more likely, though, in this study, each WS genotype was isolated from a separate microcosm to avoid selecting clones.

7.2 The genetics of adaptation.

Through the study of the mechanism of an adaptive radiation, in this study focussing on one gene and one genotype, it was hoped eventually to draw general conclusions about the genetics of adaptation. One of the perennial questions concerns the magnitude of the mutations that cause adaptive change and, more importantly, the relationship between the magnitude of the sequence change and its phenotypic consequences (Carroll, 2000; Stern, 2000). The study of the fitness of *wspR* mutants in Chapter 5 suggests that the change from SM to WS can occur due to a single mutation and that such a mutant can be fit and is not merely a 'hopeless monster'. This is consistent with studies discussed in Chapter 1 (e.g. Orr and Irving, 1997; Bull *et al.*, 2000). Had all the mutations studied in the earlier chapters diminished fitness it would have implied that this case was strictly micromutationist in that artificial alterations, with visible phenotypes, were too large to be adaptive. That this is not the case does not, however, confirm universal macromutationism as this is one specific example which has its caveats. The change does not occur in a complex multicellular organism and it could be argued that pleiotropy is less of a constraint. However, the fact that a gene such as *wspR* has various inputs and outputs, and the fact that operons are co-transcribed (which is not the case in eukaryotes) show that pleiotropy is still an obstacle that selection must overcome. Indeed Stern (2000) argues against universal pleiotropy in metazoans, while Shapiro (1997) stresses the developmental complexity of prokaryotes. Argument can also be made that the change from SM to WS is not phenotypically large and merely involves a qualitative change in CLP levels. This would put it in a similar class to mutations that confer antibiotic resistance (Schrag *et al.*, 1997; Reynolds, 2000). However, the SM to WS transition

does produce a change in environmental niche specialisation, which is the basis of adaptive radiation and, in some organisms, speciation. Hence the phenotypic consequences of this change in genotype are substantial, and the fact that it is due to a relatively simple mechanistic change gives an insight into the workings of selection, and the relationship between genetics and ecology. Further work to uncover the extent of the pleiotropy of *wspR* in terms of motility would therefore contribute to both evolutionary and molecular biology.

As well as the magnitude of the mutations responsible for adaptive evolution, it is of interest to know in which genes they are found and what effects they have on them. As discussed in Chapter 1 they are expected to occur in regulatory genes more often than structural ones, and are also more likely to be in non-coding than coding regions (Doebley and Lukens, 1998; Carroll, 2000; Purugganan, 2000). In LSWS the mutation does not occur in the structural operon (*wss*) so it was decided to look at the regulatory operon (*wsp*). Indeed the phenotype of JB01 shows that even the non-coding region of the structural operon is not a candidate for adaptive change. The fitness effects of mutations in the coding sequence of *wspR* imply that coding regulatory regions are good candidates for change, even though none were found in Chapter 6. They may, however, be found in the other genes of the operon. This is in contrast to the views of Doebley and Lukens (1998), and Carroll (2000), who suggest that non-coding regions are the most likely sites for adaptive mutations. The case of the plastic genotypes indicated the difficulty surrounding the selection of non-coding region mutations in bacteria. An increase in *wspR* transcription was shown to be beneficial but did not arise. This may be because it cannot happen without increasing the levels of transcription in the whole operon. This is especially important if the genes are also translationally coupled so that individual amounts cannot be altered at this level either (e.g. by changes in ribosome binding site sequence). This could show that in bacteria, where genes are not regulated individually, non-coding regions are of less importance for adaptive evolution than they are in eukaryotes. This would be a further example of a eukaryotic-prokaryotic difference that is of evolutionary importantance, to add to the list of Levin and Bergstrom (2000).

The alleles *wspR14* and *wspR19* that caused the SM to WS transition when overexpressed, were both dominant, gain-of-function alleles. However, they were not found to occur naturally. The reason for this remains highly speculative, but it might be that loss-of-function mutations are more common. In the *wsp* pathway the methylesterase (WspF) removes methyl groups from the MCP (WspA) thus preventing continuous activation of the histidine kinase (WspE). This is the negative element in the pathway and its absence would result in continuous activity, and possibly the WS phenotype. It is imagined that the majority of alterations to the *wspF* sequence would result in loss of function whereas the majority of mutations in the other *wsp* genes would not lead to the requisite gain of function. Hence the absence of mutations in *wspR* in WS genotypes might indicate the importance of loss of function. The *wsp* pathway is not unique in this: most metabolic and signalling pathways have a balance of positive and negative elements (reviewed in Stock *et al.*, 2000) and potentially a loss-of-function mutation could be selected to achieve both up and down-regulation. Hence loss-of-function mutations would form the basis of most short-term adaptation.

The candidate gene approach to *wspR*, used in this study, has proved successful in several respects. As discussed above, it has given insights into the types of mutations that can occur in adaptation. The approach was successful despite the failure to discover mutations in *wspR* in microcosm isolates. Much was learned about the WS phenotype and its relation to *wspR* such as the existence of constraints and adaptive peaks. The study highlighted the importance of regulatory genes, such as *wspR*, which in this case promote a primitive kind of multicellularity (c.f. Shapiro, 1998). This demonstrates the link between such prokaryotic studies and the studies of adaptations in metazoan body plans that address the questions raised by Stern (2000).

8.3 Molecular microbiology

As well as the evolutionary questions this study has addressed some mechanistic issues which have an interest of their own. GGDEF proteins are a relatively recently discovered family which are being implicated in a wider range of processes with each genome project that is completed (Galperin *et al.*, 2001). This study has uncovered several features of WspR, as well as GGDEF proteins in general through the study of PleD. The role of phosphorylation and salt input into WspR has been confirmed as well as its position as a rate-limiting step in the *wsp* pathway. On the output side its roles in CLP production, biofilm formation, attachment and motility have been demonstrated. The range of different processes involved implies the action of a common intermediate which is thought to be cyclic di-GMP. All these factors are components of the fitness effect. The study of the interchangeability of WspR and PleD implies a common mechanism for GGDEF protein action, which has also been suggested by other authors (e.g. Ausmees *et al.*, 2001), and again cyclic di-GMP looks the most likely candidate as it is a potential second messenger that could affect diverse processes. Most interesting in this study has been the fact that evolutionary and molecular insights reinforce each other, making the analysis more powerful. For example, the fitness assays in Chapter 5 led to mechanistic inferences about the phosphorylation mutants, while the genetics of the *wspR* alleles led to insights about their evolutionary consequences.

8.4 Future directions

This study has employed purely genetic means to assess the function of WspR. It could be argued that such an approach has reached its limit and there is need for a more biochemical approach. If WspR can be purified, most probably by the use of a His-tagged vector, then *in vitro* studies of both input and output can be undertaken. Given the role of phosphorylation a more extensive study of the kinetics of the process could be carried out. Also assays for the binding of cyclic di-GMP would confirm this as the crucial signalling molecule. Purified protein could also be used to raise an antibody to *wspR*, which would be useful for Western analysis of protein levels in different genotypes.

In Chapter 7 the proteomic approach to the WS phenotype was vindicated by the discovery of differences between genotypes. This has been further shown in studies that distinguish up to 1000 proteins on a gel with numerous putative differences between SM and WS (C. Knight and P.B. Rainey, personal communication). The use of mass spectrometry to identify these proteins is the obvious next step. The approach could then be applied to more of the 26 independently isolated WS genotypes to analyse the different types of WS phenotype. Proteomics is not the only post-genomic technology that could be applied. The current sequencing of the *P. fluorescens* SBW25 genome will allow for the use of microarray transcriptome analysis, which might bridge the gap between the mutations causing the WS phenotype and the effects at the proteomic level. It is conceivable that a single mutation affects the transcription of a certain set of genes that affects the translation and activity of a wider set. Another possible technology is the use of metabolome profiling which could highlight changes in small molecule concentrations, such as cyclic di-GMP, that occur with the SM to WS transition.

At an evolutionary level the combination of a sequenced genome and further mutational analysis of WS genotypes could lead to further candidate gene studies. A mutant screen, such as that carried out by Spiers *et al.* (2002), will be a significantly easier process when a sequenced genome removes the need for extensive sequencing and mapping. An analysis of the WS genotypes that differ from LSWS, both from the results in Chapter 7 and from further proteomic work, might indicate further pathways that can lead to adaptive biofilm formation. It would be of great interest to know whether these different WS genotypes co-exist in the same microcosm due to frequency-dependent selection, or dominate different microcosms due to chance mutational events. Likewise an assessment of the Fuzzy Spreader phenotype could also be made. With these kinds of experiments a complete unravelling of this adaptive radiation can occur at the molecular level, of which this study is the beginning.

Bibliography

Abrams, P.A. (1987) Alternative models of character displacement and niche shift. 2. Displacement when there is competition for a single resource. *American Naturalist* **130** (2): pp 271-282.

Aizawa, S. I., Harwood, C. S. & Kadner, R. J. (2000) Signalling components in bacterial locomotion and sensory reception. *Journal of Bacteriology* **182**: pp 1459–1471.

Akam, M. (1998) *Hox* genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *International Journal of Developmental Biology* **42** (3): pp 445-451.

Aldridge, P. & Jenal U. (1999) Cell cycle-dependent degradation of a flagellar motor component requires a novel-type response regulator. *Molecular Microbiology* **32**: pp 379–391.

Alexandre, G. & Zhulin, I.B. (2001) More Than One Way To Sense Chemicals. *Journal of Bacteriology* **183** (16): pp 4681-4686.

Allison D.G., Ruiz, B., SanJose, C., Jaspe, A. & Gilbert, P. (1998) Extracellular products as mediators of the formation and detachment of *Pseudomonas fluorescens* biofilms. *FEMS Microbiology Letters* **167** (2): pp 179-184.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215** (3): pp 403-410.

Ames, P. & Parkinson, J.S. (1994) Constitutively signalling fragments of Tsr, the *Escherichia coli* serine chemoreceptor. *Journal of Bacteriology* **176** (20): pp 6340-6348.

Amikan, D., & Benziman, M. (1989) Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **171**: pp 6649-6655.

Amor, Y., Mayer, R., Benziman, M. & Delmer, D. (1991) Evidence for a cyclic diguanylic acid-dependent cellulose synthase in plants. *Plant Cell* **3** (9): pp 989-995.

Appleby, J.L. & Bourret R.B. (1999) Activation of CheY mutant D57N by phosphorylation at an alternative site, Ser-56. *Molecular Microbiology* **34** (5): pp 915-925.

Ausmees, N., Jonsson, H., Hoglund, S., Ljunggren, H. & Lindberg, M. (1999) Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii*. *Microbiology* **145**: pp 1253-1262.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. & Lindberg, M. (2001) Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiology Letters* **204** (1): pp 163-167.

Avery, O.T., Macleod, C.M. & McCarty M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction on transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *Journal of Experimental Medicine* **79**: pp 137-158.

Bantinaki, E. (2001) Characterisation of a novel chemosensory pathway underlying adaptive evolution in experimental populations of *Pseudomonas fluorescens* SBW25. D.Phil Thesis, University of Oxford.

Barak, R. & Eisenbach, M. (1999) Chemotactic-like response of *Escherichia coli* cells lacking the known chemotaxis machinery but containing overexpressed CheY. *Molecular Microbiology* **31** (4): pp 1125-1137.

Barrier, M., Robichaux, R.H. & Purugganan, M.D. (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of The National Academy of Sciences of The United States of America* **98** (18): pp 10208-10213.

Bateson, W. (1909) Mendel's Principles of Heredity. Cambridge University Press, Cambridge, UK.

Beadle, G.W. & Tatum, E.L. (1941) Genetic control of biochemical reactions in *Neurospora*. *Proceedings of The National Academy of Sciences of The United States of America* **27**: pp 499-506.

Belting, H.G., Shashikant, C.S. & Ruddle, F.H. (1998) Modification of expression and *cis*-regulation of Hoxc8 in the evolution of diverged axial morphology. *Proceedings of The National Academy of Sciences of The United States of America* **95** (5): pp 2355-2360.

Bilwes, A.M., Alex, L.A., Crane, B.R. & Simon, M.I. (1999) Structure of CheA, a signal-transducing histidine kinase. *Cell* **96** (1): pp 131-141.

Bock, A., Bantscheff, M., Perraud, A.L., Rippe, K., Weiss, V., Glocker, M.O. & Gross, R. (2001) Rational design and molecular characterisation of a chimaeric response regulator protein. *Journal of Molecular Biology* **310** (2): pp 283-290.

Bourret, R.B., Hess, J.F. & Simon, M.I. (1990) Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. *Proceedings of The National Academy of Sciences of The United States of America* **87** (1): pp 41-45.

Boyd, J. M. (2000) Localisation of the histidine kinase PilS to the poles of *Pseudomonas aeruginosa* and identification of a localisation domain. *Molecular Microbiology* **36**: pp 153–162.

Brown, P.K., Dozois, C.M., Nickerson, C.A., Zuppardo, A., Terlonge, J. & Curtiss, R. (2001) MlrA, a novel regulator of curli (Agf) and extracellular matrix synthesis by *Escherichia coli* and *Salmonella enterica* serovar *Typhimurium. Molecular Microbiology* **41** (2): pp 349-363.

Brown, R.M., Jr. (1996) The biosynthesis of cellulose. *Journal of Macromolecular Science - Pure and Applied Chemistry* **10**: pp 1345–1373.

Buckling, A., Kassen, R., Bell, G. & Rainey, P.B. (2000) Disturbance and diversity in experimental microcosms. *Nature* **408**: pp 961-964.

Bull, J.J., Badgett, M.R. & Wichman, H.A. (2000) Big-benefit mutations in a bacteriophage inhibited with heat. *Molecular Biology and Evolution* **17** (6): pp 942-950.

Burch C.L. & Chao L. (1999) Evolution by small steps and rugged landscapes in the RNA virus phi 6. *Genetics* **151** (3): pp 921-927.

Burton, G.J., Hecht, G.B. & Newton, A. (1997) Roles of histidine protein kinase Plec in *Caulobacter crescentus* motility and chemotaxis. *Journal of Bacteriology* **179** (18): pp 5849-5853.

Carroll, S.B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**: pp 577-580.

Casaz, P., Happel, A., Keithan, J., Read, D.L., Strain, S.R. & Levy, S.B. (2001) The *Pseudomonas fluorescens* transcription activator AdnA is required for adhesion and motility. *Microbiology* **147** (2): pp 355-361.

Cavener, D.R. (1992) Transgenic animal studies on the evolution of genetic regulatory circuitries. *Bioessays* **14** (4): pp 237-244.

Charlesworth, B. (1994) The genetics of adaptation - lessons from mimicry. *American Naturalist* **144** (5): pp 839-847.

Cooper, V.S. & Lenski, R.E. (2000) The population genetics of ecological specialisation in evolving *Escherichia coli* populations. *Nature* **407**: pp 736-739.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R. & Lappinscott, H.M. (1995) Microbial Biofilms. *Annual Review of Microbiology* **49**: pp 711-745.

Crick, F.H.C., Barnett, L., Brenner, S. & Watts-Tobin, R.J. (1961) General nature of the genetic code for proteins. *Nature* **192**: pp 1227-1232.

Darwin, C. (1859) *On the Origin of Species*. Murray, London, UK.

Davey, M.E. & O'Toole, G.A. (2000) Microbial biofilms: from ecology to molecular genetics. *Microbiology and Molecular Biology Reviews* **64** (4): pp 847-869.

Davies, D.G. & Geesey, G.G. (1995) Regulation of the alginate biosynthesis gene *algC* in *Pseudomonas aeruginosa* during biofilm development in continuous culture. *Applied Environmental Microbiology* **61**: pp 860–867.

De Jong, G. (1995) Phenotypic plasticity as a product of selection in a variable environment. *American Naturalist* **145** (4): pp 493-512.

Deziel, E., Comeau, Y. & Villemur, R. (2001) Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *Journal of Bacteriology* **183** (4): pp 1195-1204.

Dickinson, G.M. & Bisno, A.L. (1993) Infections associated with prosthetic devices - clinical considerations. *International Journal of Artificial Organs* **16** (11): pp 749-754.

Dickinson, W.J., Rowan, R.G. & Brennan, M.D. (1984) Regulatory gene evolution - adaptive differences in expression of alcohol-dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. *Heredity* **52**: pp 215-221.

Ditta, G., Stanfield, S., Corbin, D. & Helinski, D.R. (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America* **77** (12): pp 7347-7351.

Doebley, J. & Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *The Plant Cell* **10**: pp 1075-1082.

Dorel, C., Vidal, O., Prigent-Combaret, C., Vallet, I. & Lejeune, P. (1999) Involvement of the Cpx signal transduction pathway of *E. coli* in biofilm formation. *FEMS Microbiology Letters* **178** (1): pp 169-175.

Dutt, M.J. & Lee, K.H. (2000) Proteomic analysis. *Current Opinion In Biotechnology* **11** (2): pp 176-179.

Edwards, K.J., Bond, P.L., Gihring, T.M. & Banfield, J.F. (2000) An archaeal iron-oxidising extreme acidophile important in acid mine drainage. *Science* **287** (5459): pp 1796-1799.

Egger, L.A., Park, H. & Inouye, M. (1997) Signal transduction via the histidyl-aspartyl phosphorelay. *Genes to Cells* **2** (3): pp 167-184.

Falke, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A. & Danielson, M.A. (1997) The two-component signalling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual Review of Cell and Developmental Biology* **13**: pp 457–512.

Feder, M.E. (1999) Engineering candidate genes in studies of adaptation: The heat-shock protein Hsp70 in *Drosophila melanogaster*. *American Naturalist* **154** (suppl.): pp S55-S66.

Feder, M.E. & Krebs, R.A. (1998) Natural and genetic engineering of the heat-shock protein Hsp70 in *Drosophila melanogaster*: Consequences for thermotolerance. *American Zoologist* **38** (3): pp 503-517.

Fellay, R., Krisch, H.M., Prentki, P. & Frey, J. (1989) Omegon-Km - a transposable element designed for *in vivo* insertional mutagenesis and cloning of genes in gram-negative bacteria. *Gene* **76** (2): pp 215-226.

Fenyo, D. (2000) Identifying the proteome: software tools. *Current Opinion in Biotechnology* **11** (4): pp 391-395.

Fischer, D. (2000) Hybrid fold recognition: combining sequence derived properties with evolutionary information. *Pacific Symposium on Biocomputing 2000*: pp 119-130.

Fisher, R.A. (1930) *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, UK.

Fry, J.D. (1996) The evolution of host specialisation: Are trade-offs overrated? *American Naturalist* **148** (suppl.): pp S84-S107.

Futuyma, D.J. & Moreno, G. (1988) The evolution of ecological specialisation. *Annual Review of Ecology and Systematics* **19**: pp 207-233.

Galperin, M.Y. (2001) Conserved 'hypothetical' proteins: new hints and new puzzles. *Comparative and Functional Genomics* **2** (1): pp 14-18.

Galperin, M.Y., Nikolskaya, A.N. & Koonin, E.V. (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiology Letters* **203** (1): pp 11-21.

Golding, G.B. & Dean, A.M. (1998) The structural basis of molecular adaptation. *Molecular Biology and Evolution* **15** (4): pp 355-369.

Goldschmidt, R.B. (1940) *The Material Basis of Evolution.* Yale University Press, New Haven, CT.

Gould, S.J. (1977) *Ontogeny and Phylogeny.* Belknap Press, Cambridge, MA.

Gregorius, H.R. (2000) Adaptation to spatially heterogeneous modifying and adaptive environments. *Journal of Theoretical Biology* **205** (1): pp 121-131.

Grimson, M.J., Haigler, C.H. & Blanton, R.L. (1996) Cellulose micro-fibrils, cell motility, and plasma membrane protein organisation change in parallel during culmination in *Dictyostelium discoideum*. *Journal of Cell Science* **109**: pp 3079–3087.

Grunenfelder, B., Rummel, G., Vohradsky, J., Roder, D., Langen, H. & Jenal, U. (2001) Proteomic analysis of the bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America* **98**: pp 4681–4686.

Gygi, S.P. & Aebersold, R. (2000) Mass spectrometry and proteomics. *Current Opinion in Chemical Biology* **4** (5): pp 489-494.

Gygi, S., Rochon, Y., Franza, B. & Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Molecular Cell Biology* **19**: pp 1720-1730.

Haag, E.S. & True, J.R. (2001) Perspective: From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. *Evolution* **55** (6): pp 1077-1084.

Han. B., Pain, A. & Johnstone, K. (1997) Spontaneous duplication of a 661 bp element within a two-component sensor regulator gene causes phenotypic switching in colonies of *Pseudomonas tolaasii*, cause of brown blotch disease of mushrooms. *Molecular Microbiology* **25** (2): pp 211-218.

Hanna, S.L., Sherman, N.E., Kinter, M.T. & Goldberg, J.B. (2000) Comparison of proteins expressed by *Pseudomonas aeruginosa* strains representing initial and chronic isolates from a cystic fibrosis patient: an analysis by 2-D gel electrophoresis and capillary column liquid chromatography-tandem mass spectrometry. *Microbiology* **146** (10): pp 2495-2508.

Harrison, D.M., Skidmore, J., Armitage, J.P. & Maddock, J.R. (1999) Localisation and environmental regulation of MCP-like proteins in *Rhodobacter sphaeroides*. *Molecular Microbiology* **31**: pp 885–892.

Hecht, G.B. & Newton, A. (1995) Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. *Journal of Bacteriology* **177**: pp 6223–6229.

Hegner, R.E. & Wingfield, J.C. (1987) Effects of experimental manipulation of testosterone levels on parental investment and breeding success in male house sparrows. *Auk* **104** (3): pp 462-469.

Herrero, M., De Lorenzo, V. & Timmis, K.N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *Journal of Bacteriology* **172** (11): pp 6557-6567.

Hoch, J.A. (1993) Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Annual Review of Microbiology* **47**: pp 441-465.

Hoch, J.A. (2000) Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology* **3**: pp 165–170.

Ishige, K., Nagasawa, S., Tokishita, S. & Mizuno, T. (1994) A novel device of bacterial signal transducers. *EMBO Journal* **13** (21): pp 5195-5202.

Jenal, U. (2000) Signal transduction mechanisms in *Caulobacter crescentus* development and cell cycle control. *FEMS Microbiology Reviews* **24**: pp 177–191.

Jones, H.A., Lillard, J.W. & Perry, R.D. (1999) HmsT, a protein essential for expression of the haemin storage (Hms+) phenotype of *Yersinia pestis*. *Microbiology* **145** (8): pp 2117-2128.

Kahn, S.G. (1998) Molecular characterisation of genes essential for adaptive evolution in *Pseudomonas fluorescens* SBW25. D.Phil Thesis, University of Oxford.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. & Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* **7** (6): pp 331-338.

Kassen, R., Buckling, A., Bell, G. & Rainey, P.B. (2000) Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406** (6795): pp 508-512.

Kearns, D.B. & Shimkets, L.J. (2001) Lipid chemotaxis and signal transduction in *Myxococcus xanthus*. *Trends in Microbiology* **9** (3): pp 126-129.

Ketterson, E.D. & Nolan, V. (1999) Adaptation, exaptation, and constraint: A hormonal perspective. *American Naturalist* **154** (suppl.): pp S4-S25.

Kim, C., Jackson, M., Lux, R. & Khan, S. (2001) Determinants of chemotactic signal amplification in *Escherichia coli*. *Journal of Molecular Biology* **307** (1): pp 119-135.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura, S., Chen, H.P., Saxena, I.M., Brown, R.M., Jr. & Itoh, T. (2001) Localisation of c-di-GMP-binding protein with the linear terminal complexes of *Acetobacter xylinum*. *Journal of Bacteriology*, **183** (19): pp 5668–5674.

Kirschner, M. & Gerhart, J. (1998) Evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **95**: pp 8420-8427.

Korber, D.R., Lawrence, J.R. & Caldwell, D.E. (1994) Effect of motility on surface colonisation and reproductive success of *Pseudomonas fluorescens* in dual-dilution continuous culture and batch culture systems. *Applied and Environmental Microbiology* **60**: pp 1421–1429.

Krebs, R.A. & Feder, M.E. (1998) Experimental manipulation of the cost of thermal acclimation in *Drosophila melanogaster*. *Biological Journal of the Linnean Society* **63** (4): pp 593-601.

Lack, D. (1947) *Darwin's Finches*. Cambridge University Press, Cambridge, UK.

Lamb, J.R., Tugendreich, S. & Hieter, P. (1995) Tetratrico peptide repeat interactions - to TPR or not to TPR. *Trends in Biochemical Sciences* **20** (7): pp 257-259.

Larsen, M.R., Sorensen, G.L., Mose-Larsen, P., Fey, S.J. & Roepstorff, P. (2001) Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis. *Proteomics* **1** (2): pp 223-228.

Lenski, R.E., Rose, M.R., Simpson, S.C. & Tadler, S.C. (1991) Long-term experimental evolution in *Escherichia coli* .1. Adaptation and divergence during 2,000 generations. *American Naturalist 138* (6): pp 1315-1341.

Levin, B.R. & Bergstrom, C.T. (2000) Bacteria are different: Observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **97** (13): pp 6981-6985.

Levin, B.R., Perrot, V. & Walker, N. (2000) Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **154** (3): pp 985-997.

Loper, J.E. & Lindow, S.E. (1994) A biological sensor for iron available to bacteria in their habitats on plant-surfaces. *Applied and Environmental Microbiology* **60** (6): pp 1934-1941.

Lybarger, S.R. & Maddock, J.R. (2001) Polarity in action: asymmetric protein localisation in bacteria. *Journal of Bacteriology* **183** (11): pp 3261-3267.

Maddock, J.R. & Shapiro, L. (1993) Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259** (5102): pp 1717-1723.

Matthysse, A.G. (1983) Role of bacterial cellulose fibrils in *Agrobacterium tumefaciens* infection. *Journal of Bacteriology* **154** (2): pp 906-915.

Matthysse, A.G., White, S. & Lightfoot, R. (1995a) Genes required for cellulose synthesis in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **177**: pp 1069-1075.

Matthysse, A.G., Thomas, D.L., & White, A.R. (1995b) Mechanism of cellulose synthesis in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **177**: pp 1076-1081.

Mayr, E., & Provine, W.B. (1980) *The evolutionary synthesis. Perspectives on the unification of biology.* Harvard University Press, Cambridge, MA.

Miller, J.H. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Morrison, T.B. & Parkinson, J.S. (1994) Liberation of an interaction domain from the phosphotransfer region of CheA, a signalling kinase of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **91** (12): pp 5485-5489.

Nakashima, K., Kanamaru, K., Aiba, H. & Mizuno, T. (1991) Osmoregulatory expression of the porin genes in *Escherichia coli* - evidence for signal titration in the signal transduction through EnvZ-OmpR phosphotransfer. *FEMS Microbiology Letters* **82** (1): pp 43-47.

Olsen, A., Jonsson, A. & Normark, S. (1989) Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. *Nature* **338** (6217): pp 652-655.

Orr, H.A. (1999) The evolutionary genetics of adaptation: a simulation study. *Genetical Research* **74** (3): pp 207-214.

Orr, H.A. & Coyne, J.A. (1992) The genetics of adaptation - a reassessment. *American Naturalist* **140** (5): pp 725-742.

Orr, H.A. & Irving, S. (1997) The genetics of adaptation: the genetic basis of resistance to wasp parasitism in *Drosophila melanogaster*. *Evolution* **51** (6): pp 1877-1885.

O'Toole, G.A. & Kolter, R. (1998a) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Molecular Microbiology* **30**: pp 295–304.

O'Toole, G.A. & Kolter, R. (1998b) Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Molecular Microbiology* **28** (3): pp 449-461.

O'Toole, G.A., Pratt, P.A., Watnick, P.I., Newman, D.K., Weaver, V.B. & Kolter, R. (1999) Genetic approaches to study biofilms. *Methods in Enzymology* **310**: pp 91–109.

O'Toole, G., Kaplan, H.B. & Kolter, R (2000) Biofilm formation as microbial development. *Annual Review of Microbiology* **54**: pp 49–79.

Overmann, J. & Pfennig, N. (1992) Buoyancy regulation and aggregate formation in *Amoebobacter purpureus* from Mahoney Lake. *FEMS Microbiology Ecology* **101** (2): pp 67-79.

Palopoli, M.F. & Patel, N.H. (1996) Neo-Darwinian developmental evolution: Can we bridge the gap between pattern and process? *Current Opinion in Genetics and Development* **6** (4): pp 502-508.

Palopoli, M.F. & Patel N.H. (1998) Evolution of the interaction between *hox* genes and a downstream target. *Current Biology* **8** (10): pp 587-590.

Pao, G.M. & Saier, M.H., Jr. (1995) Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *Journal of Molecular Evolution* **40**: pp 136–154.

Paquin, C. & Adams, J. (1983) Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature* **302** (5908): pp 495-500.

Pei, J. & Grishin , N.V. (2001) GGDEF domain is homologous to Adenylyl Cyclase. *Proteins: Structure, Function, and Genetics* **42**: pp 210–216.

Pendrak, M.L. & Perry, R.D. (1991) Characterisation of a haemin-storage locus of *Yersinia pestis*. *Biology of Metals* **4** (1): 41-47.

Perego, M. (1998) Kinase-phosphatase competition regulates *Bacillus subtilis* development. *Trends in Microbiology* **6** (9): pp 366-370.

Pratt, L.A. & Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Molecular Microbiology* **30**: pp 285–293.

Pratt, L.A. & Kolter, R. (1999) Genetic analyses of bacterial biofilm formation. *Current Opinion in Microbiology* **2** (6): pp 598-603.

Provine, W.B. (1971) *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago, IL.

Ptashne, M. (1992) *A Genetic Switch: Phage λ and Higher Organisms*. Cell Press, Cambridge, MA.

Purugganan, M.D. (1998) The molecular evolution of development. *Bioessays* **20** (9): pp 700-711.

Purugganan, M.D. (2000) The molecular population genetics of regulatory genes. *Molecular Ecology* **9** (10): pp 1451-1461.

Purugganan, M.D. & Suddith, J.I. (1998) Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: Nonneutral evolution and naturally occurring variation in floral homeotic function. *Proceedings of the National Academy of Sciences of the United States of America* **95** (14): pp 8130-8134.

Rainey, P.B. & Bailey, M.J. (1996) Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Molecular Microbiology* **19** (3): pp 521-533.

Rainey, P.B., Buckling, A., Kassen, R. & Travisano, M. (2000) The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends in Ecology and Evolution* **15** (6): pp 243-247.

Rainey, P.B. & Travisano, M. (1998) Adaptive radiation in a heterogeneous environment. *Nature* **394** (6688): pp 69-72.

Rashid, M.H., Rumbaugh, K., Passador, L., Davies, D.G., Hamood, A.N., Iglewski, B.H. & Kornberg, A. (2000) Polyphosphate kinase is essential for biofilm development, quorum sensing, and virulence of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America* **97** (17): pp 9636-9641.

Reynolds, M.G. (2000) Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* **156** (4): pp 1471-1481.

Romling, U., Rohde, M., Olsen, A., Normark, S. & Reinkoster, J. (2000) AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. *Molecular Microbiology* **36** (1): pp 10-23.

Rosenweig, R.F., Sharp, R.R., Treves, D.S. & Adams, J. (1994) Microbial evolution in a simple unstructured environment - genetic differentiation in *Escherichia coli*. *Genetics* **137** (4): pp 903-917.

Ross, P., Mayer, R., Weinhouse, H., Amikam, D., Huggirat, Y., Benziman, M., de Vroom, E., Fidder, A., de Paus, P. & Sliedregt, L.A. (1990) The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinu*m. Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *Journal of Biological Chemistry* **265**: pp 18933–18943.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Ohana, P., Mayer, R., Braun, S., De Vroom, E., Van Der Marel, G.A., Van Boom, J.H. & Benziman M. (1987) Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**: pp 279-281.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd edition.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Sanders, D.A., Gillececastro, B.L., Stock, A.M., Burlingame, A.L. & Koshland, D.E. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *Journal of Biological Chemistry* **264** (36): pp 21770-21778.

Saxena, I.M., Kudlicka, K., Okuda, K. & Brown, R.M., Jr. (1994) Characterisation of genes in the cellulose-synthesising operon *(acs* operon) of *Acetobacter xylinu*m: implications for cellulose crystallization. *Journal of Bacteriology* **176**: pp 5735–5752.

Scheiner, S.M. (1993) Plasticity as a selectable trait – reply. *American Naturalist* **142** (2): pp 371-373.

Schlichting, C.D. & Pigliucci, M. (1993) Control of phenotypic plasticity via regulatory genes. *American Naturalist* **142** (2): pp 366-370.

Schlosser, A., Pipkorn, R., Bossemeyer, D. & Lehmann, W.D. (2001) Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Analytical Chemistry* **73** (2): pp 170-176.

Schluter, D. (1994) Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266** (5186): pp 798-801.

Schmidt, P.S., Duvernell, D.D. & Eanes, W.F. (2000) Adaptive evolution of a candidate gene for ageing in *Drosophila. Proceedings of the National Academy of Sciences of the United States of America* **97** (20): pp 10861-10865.

Schmitt, J. (1999) Introduction: Experimental approaches to testing adaptation. *American Naturalist* **154** (suppl.): pp S1-S3.

Schrag, S.J., Perrot, V. & Levin, B.R. (1997) Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli. Proceedings of the Royal Society of London Series B – Biological Sciences* **264** (1386): pp 1287-1291.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. University of Illinois Press, Urbana, IL.

Shapiro, J.A. (1997) Genome organisation, natural genetic engineering and adaptive mutation. *Trends in Genetics* **13** (3): pp 98-104.

Shapiro, J.A. (1998) Thinking about bacterial populations as multicellular organisms. *Annual Review of Microbiology* **52**: pp 81-104.

Simms, S.A., Keane, M.G. & Stock, J. (1985) Multiple forms of the CheB methylesterase in bacterial chemosensing. *Journal of Biological Chemistry* **260** (18): pp 161-168.

Simon, R., Priefer, U. & Puhler, A. (1983) A broad host range mobilisation system for *in vivo* genetic engineering - transposon mutagenesis in gram-negative bacteria. *Biotechnology* **1** (9): 784-791.

Sinervo, B. (1999) Mechanistic analysis of natural selection and a refinement of Lack's and Williams's principles. *American Naturalist* **154** (suppl.): pp S26-S42.

Sommer, J.M. & Newton, A. (1989) Turning off flagellum rotation requires the pleiotropic gene *pleD* - *pleA*, *pleC*, and *pleD* define 2 morphogenic pathways in *Caulobacter crescentus*. *Journal of Bacteriology* **171** (1): pp 392-401.

Spiers, A.J., Buckling, A. & Rainey, P.B. (2000) The causes of *Pseudomonas* diversity. *Microbiology* **146** (10): pp 2345-2350.

Spiers, A.J., Kahn, S.G., Bohannon, J., Travisano, M. & Rainey, P.B. (2002) Adaptive divergence in experimental populations of Pseudomonas fluorescens. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* in press.

Stern, D.L. (2000) Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54** (4): pp 1079-1091.

Stock, A.M., Mottonen, J.M., Stock, J.B. & Schutt, C.E. (1989) 3-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature* **337** (6209): pp 745-749.

Stock, A.M., Robinson, V.L. & Goudreau, P.N. (2000) Two-component signal transduction. *Annual Review of Biochemistry* **69**: pp 183–215.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K.S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. & Olson, M.V. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406** (6799): pp 959-964.

Swissa, M., Aloni, Y., Weinhouse, H. & Benziman, M. (1980) Intermediary steps in *Acetobacter xylinum* cellulose synthesis: studies with whole cells and cell-free preparations of wild type and cellulose-less mutant. *Journal of Bacteriology* **143**: pp 1142-1150.

Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P. & Benziman, M. (1998) Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organisation and occurrence of conserved domains in isoenzymes. *Journal of Bacteriology* **180**: pp 4416-4425.

Tanaka, T., Saha, S.K., Tomomori, C., Ishima, R., Liu, D.J., Tong, K.I., Park, H., Dutta, R., Qin, L., Swindells, M.B., Yamazaki, T., Ono, A.M., Kainosho, M., Inouye, M., Ikura, M. (1998) NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* **396** (6706): pp 88-92.

Tatar, M. (1999) Transgenes in the analysis of life span and fitness. *American Naturalist* **154** (suppl.): pp S67-S81.

Tautz, D. (2000) Evolution of transcriptional regulation. *Current Opinion in Genetics and Development* **10** (5): pp 575-579.

Tautz, D. & Schmid, K.J. (1998) From genes to individuals: developmental genes and the generation of the phenotype. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* **353** (1366): pp 231-240.

Tobisch, S., Zuhlke, D., Bernhardt, J., Stulke, J. & Hecker, M. (1999) Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis*. *Journal of Bacteriology* **181**: pp 6996-7004.

Tomomori, C., Tanaka, T., Dutta, R., Park, H.Y., Saha, S.K., Zhu, Y., Ishima, R., Liu, D.J., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M. & Ikura, M. (1999) Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. *Nature Structural Biology* **6** (8): pp 729-734.

Tran, V.K., Oropeza, R. & Kenney, L.J. (2000) A single amino acid substitution in the C terminus of OmpR alters DNA recognition and phosphorylation. *Journal of Molecular Biology* **299** (5): pp 1257-1270.

Travisano, M. (2001) Experimental evolution studies yield insights into bacterial diversity - tracking the rise of diversity in vitro could help explain how mixtures of microbial species persist stably in nature. *ASM News* **67** (8): pp 403-409.

Travisano, M. & Rainey, P.B. (2000) Studies of adaptive radiation using model microbial systems. *American Naturalist* **156** (suppl.): pp S35-S44.

Turner, J.G. (1977) Butterfly mimicry: the genetical evolution of an adaptation. *Evolutionary Biology* **10**: pp 163-206.

Turner, P.E., Souza, V. & Lenski, R.E. (1996) Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology* 77 (7): pp 2119-2129.

Van Bogelen R.A., Schiller, E.E., Thomas, J.D. & Neidhardt, F.C. (1999) Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis* **20**: pp 2149-2159.

Via, S. (1993a) Adaptive phenotypic plasticity - target or by-product of selection in a variable environment. *American Naturalist* **142** (2): pp 352-365.

Via, S. (1993b) Regulatory genes and reaction norms. *American Naturalist* **142** (2): pp 374-378

Vidal, O., Longin, R., Prigent-Combaret, C., Dorel, C., Hooreman, M. & Lejeune, P. (1998) Isolation of an *Escherichia coli K-12* mutant strain able to form biofilms on inert surfaces: Involvement of a new *ompR* allele that increases curli expression. *Journal of Bacteriology* **180** (9): pp 2442-2449.

Watnick, P.I. & Kolter, R. (1999) Steps in the development of a *Vibrio cholerae* biofilm. *Molecular Microbiology* **34**: pp 586–595.

Watson, J.D. & Crick, F.H.C. (1953) Molecular structure of nucleic acid. A structure for deoxyribose nucleic acid. *Nature* **171**: pp 737-738.

Weinhouse, H., Sapir, S., Amikam, D., Shilo, Y., Volman, G., Ohana, P. & Benziman, M. (1997) C-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. *FEBS Letters* **416** (2): pp 207-211.

Wilkins, M.R., Gasteiger, E., Gooley, A.A., Herbert, B.R., Molloy, M.P., Binz, P.A., Ou, K., Sanchez, J.C., Bairoch, A., Williams, K.L. & Hochstrasser, D.F. (1999) High-throughput mass

spectrometric discovery of protein post-translational modifications. *Journal of Molecular Biology* **289**: pp 645-657.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., Decruz, E., Noyerweidner, M., Smith, S.S., Michael, M.Z. & Graham, M.W. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Research* **17** (9): pp 3469-3478.

Wong, H.C., Fear, A.L., Calhoon, R.D., Eichinger, G.H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D.H., Meade, J.H., Emerick, A.W., Bruner, R., Ben-Bassat, A. & Tal, R. (1990) Genetic organisation of the cellulose synthase operon in *Acetobacter xylinum*. *Proceedings of the National Academy of Sciences of the United States of America* **87**: pp 8130–8134.

Xu, G.W. & Gross, D.C. (1988) Physical and functional analyses of the *syrA* gene and *syrB* gene involved in syringomycin production by *Pseudomonas syringae* pv *syringae*. *Journal of Bacteriology* **170** (12): pp 5680-5688.

Zogaj, X., Nimtz, M., Rohde, M., Bokranz, W. & Romling, U. (2001) The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Molecular Microbiology* **39** (6): pp 1452-1463.

Appendix
(Statistical Analysis)

Chapter 3

A. One-way ANOVA of the effect of genotype on *wss* transcription (Figure 3.5.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	1841147	13	141626.7	66.26033	1.72E-17	2.088932
Within Genotypes	59848	28	2137.429			
Total	1900995	41				

B. Two-way ANOVA of the effects of genotype and environment on *wss* transcription (Figure 3.6.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Genotype	811432.5	2	405716.2	21.01477	8.97E-07	3.259444
Environment	686920.4	5	137384.1	7.116045	0.000102	2.477165
Interaction	403726.2	10	40372.62	2.091169	0.051625	2.106056
Within	695024.7	36	19306.24			
Total	2597104	53				

C. Two-way ANOVA of the effects of salt presence and treatment on *wss* transcription (Figure 3.6.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Salt	26048.07	1	26048.07	1.349205	0.253061	4.113161
Treatment	1502301	8	187787.7	9.726786	4.6E-07	2.208516
Interaction	373729.6	8	46716.2	2.419746	0.033298	2.208516
Within	695024.7	36	19306.24			
Total	2597104	53				

Chapter 5

A. GLM output of fitness of SM genotype overexpressing *wspR* alleles (divided into dominant and dominant-negative classes) competed against SM*panB* (Figure 5.1.):

```
                    The GLM Procedure
                Class Level Information
Class                Levels    Values
allel                     7    12 13 14 19 5 9 pl
group                     3    cont dom neg
env                       2    struct unstruct
rep                       3    1 2 3

                Number of observations    42
```

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2.05088870	0.22787652	10.23	<.0001
Error	32	0.71301359	0.02228167		
Corrected Total	41	2.76390229			

R-Square	Coeff Var	Root MSE	w Mean
0.742026	13.85163	0.149270	1.077638

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	0.02167620	0.01083810	0.49	0.6193
allel(group)	4	0.67374208	0.16843552	7.56	0.0002
env	1	0.39820802	0.39820802	17.87	0.0002
group*env	2	0.95726240	0.47863120	21.48	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	0.02167620	0.01083810	0.49	0.6193
allel(group)	4	0.67374208	0.16843552	7.56	0.0002
env	1	0.14152080	0.14152080	6.35	0.0169
group*env	2	0.95726240	0.47863120	21.48	<.0001

---------- env=struct ----------

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1.72624951	0.28770825	27.71	<.0001
Error	14	0.14536353	0.01038311		
Corrected Total	20	1.87161304			

R-Square	Coeff Var	Root MSE	w Mean
0.922332	8.672064	0.101898	1.175009

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	0.53694956	0.26847478	25.86	<.0001
allel(group)	4	1.18929995	0.29732499	28.64	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	0.53694956	0.26847478	25.86	<.0001
allel(group)	4	1.18929995	0.29732499	28.64	<.0001

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	0.010383
Critical Value of Dunnett's t	2.39709

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
dom - cont	0.28278	0.11994	0.44562	***
neg - cont	-0.05140	-0.21424	0.11144	

---------- env=unstruct ----------

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.45502728	0.07583788	27.19	<.0001
Error	14	0.03905396	0.00278957		
Corrected Total	20	0.49408124			

R-Square	Coeff Var	Root MSE	w Mean
0.920956	5.387958	0.052816	0.980267

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	0.44198905	0.22099452	79.22	<.0001
allel(group)	4	0.01303823	0.00325956	1.17	0.3664

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	0.44198905	0.22099452	79.22	<.0001
allel(group)	4	0.01303823	0.00325956	1.17	0.3664

---------- env=unstruct ----------

The GLM Procedure

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	0.00279
Critical Value of Dunnett's t	2.39709

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
neg - cont	-0.07827	-0.16267 0.00614	
dom - cont	-0.34725	-0.43165 -0.26285	***

B. GLM output of fitness of SM genotype overexpressing *wspR* alleles (divided into dominant and dominant-negative classes) competed against WS*panB* (Figure 5.1.):

The GLM Procedure

Class Level Information

Class	Levels	Values
allele	7	12 13 14 19 5 9 pl
class	3	cont dom neg
env	2	struct unstruct
rep	3	1 2 3

Number of observations 42

NOTE: Due to missing values, only 37 observations can be used in this analysis.

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1.91707405	0.21300823	19.51	<.0001
Error	27	0.29473669	0.01091617		
Corrected Total	36	2.21181074			

R-Square	Coeff Var	Root MSE	w Mean
0.866744	10.70193	0.104480	0.976277

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	0.15876014	0.07938007	7.27	0.0030
allele(class)	4	0.63313100	0.15828275	14.50	<.0001
env	1	0.50462415	0.50462415	46.23	<.0001
class*env	2	0.62055875	0.31027938	28.42	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class	2	0.18631634	0.09315817	8.53	0.0013
allele(class)	4	0.64085124	0.16021281	14.68	<.0001
env	1	0.58500448	0.58500448	53.59	<.0001
class*env	2	0.62055875	0.31027938	28.42	<.0001

---------- env=struct ----------

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Model	2	0.11416544	0.05708272	1.64	0.2218
Error	18	0.62671880	0.03481771		
Corrected Total	20	0.74088424			

R-Square	Coeff Var	Root MSE	w Mean
0.154093	21.42617	0.186595	0.870875

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	0.11416544	0.05708272	1.64	0.2218

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class	2	0.11416544	0.05708272	1.64	0.2218

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	0.034818
Critical Value of Dunnett's t	2.34238

Comparisons significant at the 0.05 level are indicated by ***.

class Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
dom - cont	0.08396	-0.20742 0.37534	
neg - cont	-0.07529	-0.36667 0.21610	

---------- env=unstruct ----------

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.62254646	0.31127323	13.10	0.0008
Error	13	0.30886913	0.02375916		
Corrected Total	15	0.93141559			

R-Square	Coeff Var	Root MSE	w Mean
0.668387	13.82897	0.154140	1.114617

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	0.62254646	0.31127323	13.10	0.000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class	2	0.62254646	0.31127323	13.10	0.0008

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	13
Error Mean Square	0.023759
Critical Value of Dunnett's t	2.43793

Comparisons significant at the 0.05 level are indicated by ***

class Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
neg - cont	-0.12641	-0.38573	0.13290	
dom - cont	-0.48480	-0.75052	-0.21908	***

C. GLM output of fitness of LSWS genotype overexpressing *wspR* alleles (divided into dominant and dominant-negative classes) competed against SM*panB* (Figure 5.1.):

The GLM Procedure

Class Level Information

Class	Levels	Values
allel	7	12 13 14 19 5 9 pl
group	3	cont dom neg
env	2	struct unstruct
rep	3	1 2 3

Number of observations 42

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	3.72471372	0.28651644	24.78	<.0001
Error	28	0.32370941	0.01156105		
Corrected Total	41	4.04842313			

R-Square	Coeff Var	Root MSE	w Mean
0.920041	12.24359	0.107522	0.878193

Source	DF	Type I SS	Mean Square	F Value	Pr > F
allel	6	3.31859170	0.55309862	47.84	<.0001
env	1	0.00154800	0.00154800	0.13	0.7172
allel*env	6	0.40457401	0.06742900	5.83	0.0005

Source	DF	Type III SS	Mean Square	F Value	Pr > F
allel	6	3.31859170	0.55309862	47.84	<.0001
env	1	0.00154800	0.00154800	0.13	0.7172
allel*env	6	0.40457401	0.06742900	5.83	0.0005

---------- env=struct ----------

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2.61411668	0.43568611	34.08	<.0001
Error	14	0.17896640	0.01278331		
Corrected Total	20	2.79308307			

R-Square	Coeff Var	Root MSE	w Mean
0.935925	12.78615	0.113063	0.884264

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	2.58473833	1.29236917	101.10	<.0001
allel(group)	4	0.02937834	0.00734459	0.57	0.6857

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	2.58473833	1.29236917	101.10	<.0001
allel(group)	4	0.02937834	0.00734459	0.57	0.6857

The GLM Procedure

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	0.012783
Critical Value of Dunnett's t	2.39709

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
neg - cont	-0.57849	-0.75917	-0.39781	***
dom - cont	-1.03067	-1.21136	-0.84999	***

---------- env=unstruct ----------

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1.10904904	0.18484151	17.88	<.0001
Error	14	0.14474302	0.01033879		
Corrected Total	20	1.25379206			

R-Square	Coeff Var	Root MSE	w Mean
0.884556	11.65890	0.101680	0.872122

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	1.06991442	0.53495721	51.74	<.0001
allel(group)	4	0.03913462	0.00978365	0.95	0.4663

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	1.06991442	0.53495721	51.74	<.0001
allel(group)	4	0.03913462	0.00978365	0.95	0.4663

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	0.010339
Critical Value of Dunnett's t	2.39709

Comparisons significant at the 0.05 level are indicated by ***.

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
neg - cont	-0.03343	-0.19592 0.12906	
dom - cont	-0.48065	-0.64314 -0.31816	***

D. GLM output of fitness of LSWS genotype overexpressing *wspR* alleles (divided into dominant and dominant-negative classes) competed against WS*panB* (Figure 5.1.):

The GLM Procedure

Class Level Information

Class	Levels	Values
allele	7	12 13 14 19 5 9 pl
class	3	cont dom neg
env	2	struct unstruct
rep	3	1 2 3

Number of observations 42

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2.22523925	0.24724881	20.73	<.0001
Error	32	0.38162266	0.01192571		
Corrected Total	41	2.60686191			

R-Square	Coeff Var	Root MSE	w Mean
0.853608	11.86168	0.109205	0.920653

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	1.44775046	0.72387523	60.70	<.0001
allele(class)	4	0.14006241	0.03501560	2.94	0.0356
env	1	0.53586120	0.53586120	44.93	<.0001
class*env	2	0.10156518	0.05078259	4.26	0.0229

Source	DF	Type III SS	Mean Square	F Value	Pr > F

class	2	1.44775046	0.72387523	60.70	<.0001
allele(class)	4	0.14006241	0.03501560	2.94	0.0356
env	1	0.24671326	0.24671326	20.69	<.0001
class*env	2	0.10156518	0.05078259	4.26	0.0229

---------- env=struct ----------

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.86461497	0.43230749	64.96	<.0001
Error	18	0.11978160	0.00665453		
Corrected Total	20	0.98439657			

R-Square	Coeff Var	Root MSE	w Mean
0.878320	10.09972	0.081575	0.807699

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	0.86461497	0.43230749	64.96	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class	2	0.86461497	0.43230749	64.96	<.0001

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	0.006655
Critical Value of Dunnett's t	2.34238

Comparisons significant at the 0.05 level are indicated by ***.

class Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
neg - cont	-0.24007	-0.36746 -0.11269	***
dom - cont	-0.55808	-0.68547 -0.43070	***

---------- env=unstruct ----------

Dependent Variable: w

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.68470067	0.34235033	15.33	0.0001
Error	18	0.40190347	0.02232797		
Corrected Total	20	1.08660414			

R-Square	Coeff Var	Root MSE	w Mean
0.630129	14.45671	0.149425	1.033607

Source	DF	Type I SS	Mean Square	F Value	Pr > F
class	2	0.68470067	0.34235033	15.33	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
class	2	0.68470067	0.34235033	15.33	0.0001

Dunnett's t Tests for w

NOTE: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	0.022328

```
Critical Value of Dunnett's t  2.34238

Comparisons significant at the 0.05 level are indicated by ***.

                    Difference
   class              Between      Simultaneous 95%
Comparison            Means      Confidence Limits

neg  - cont         0.05950    -0.17384  0.29284
dom  - cont        -0.31812    -0.55146 -0.08478  ***
```

E. One-way ANOVA of the effect of genotype on fitness over 7 days competed against SM*panB* (Figure 5.2.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	129.1159	7	18.44512	24.53052	3.96E-07	2.706628
Within Genotypes	11.27888	15	0.751925			
Total	140.3947	22				

F. One-way ANOVA of the effect of genotype on fitness over 7 days competed against WS*panB* (Figure 5.2.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	4.182255	7	0.597465	55.33695	4.98E-10	2.657195
Within Genotypes	0.17275	16	0.010797			
Total	4.355005	23				

G. One-way ANOVA of the effect of genotype on fitness when invading a population of SM*panB* from rare (Figure 5.3.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	110.6591	9	12.29546	15.64807	3.18E-07	2.392817
Within Genotypes	15.71498	20	0.785749			
Total	126.3741	29				

H. One-way ANOVA of the effect of genotype on fitness when invading a population of WS*panB* from rare (Figure 5.3.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	4.869455	9	0.541051	10.44429	8.24E-06	2.392817
Within Genotypes	1.03607	20	0.051803			

Total	5.905525	29				

I. One-way ANOVA of the effect of recipient genotype on fitness of SM*panB* when invading from rare (Figure 5.3.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	4.198286	9	0.466476	5.1095	0.001618	2.456282
Within Genotypes	1.643326	18	0.091296			
Total	5.841611	27				

J. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on SM fitness in a structured microcosm competed against SM*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.087864	2	0.043932	14.29985	0.005215	5.143249
Within Alleles	0.018433	6	0.003072			
Total	0.106297	8				

K. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on SM fitness in a structured microcosm competed against WS*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.129085	2	0.064543	68.65302	7.34E-05	5.143249
Within Alleles	0.005641	6	0.00094			
Total	0.134726	8				

L. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on LSWS fitness in a structured microcosm competed against SM*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.096417	2	0.048208	5.697207	0.041042	5.143249
Within Alleles	0.050771	6	0.008462			
Total	0.147187	8				

M. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on LSWS fitness in a structured microcosm competed against WS*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*

Between Alleles	0.125179	2	0.06259	72.81963	0.000201	5.786148
Within Alleles	0.004298	5	0.00086			
Total	0.129477	7				

N. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on SM fitness in an unstructured microcosm competed against SM*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.016161	2	0.008081	5.902117	0.048292	5.786148
Within Alleles	0.006845	5	0.001369			
Total	0.023007	7				

O. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on SM fitness in an unstructured microcosm competed against WS*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.036128	2	0.018064	1.380548	0.375772	9.552082
Within Alleles	0.039253	3	0.013084			
Total	0.075381	5				

P. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on LSWS fitness in an unstructured microcosm competed against SM*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.31948	2	0.15974	21.54366	0.001826	5.143249
Within Alleles	0.044488	6	0.007415			
Total	0.363969	8				

Q. One-way ANOVA of the effect of *wspR* phosphorylation-deficient allele on LSWS fitness in an unstructured microcosm competed against WS*panB* (Figure 5.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Alleles	0.113585	2	0.056793	65.72098	8.32E-05	5.143249
Within Alleles	0.005185	6	0.000864			
Total	0.11877	8				

Chapter 7

A. Two-way ANOVA of the effects of WS genotype and *wspR* allele on *wss* transcription (Figure 7.4.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Allele	5629017	3	1876339	74.92975	3.97E-18	2.79806
Genotype	1213334	5	242666.9	9.690663	1.92E-06	2.408513
Interaction	2617153	15	174476.9	6.967562	1.1E-07	1.880174
Within	1201983	48	25041.31			
Total	10661487	71				

B. Two-way ANOVA of the effects of WS genotype and *wspR* allele on attachment (Figure 7.5.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Genotype	384736.1	5	76947.21	3.57639	0.004746	2.289852
Allele	3075594	3	1025198	47.64966	2.36E-20	2.680167
Interaction	2227364	15	148490.9	6.901634	1.35E-10	1.750497
Within	2581839	120	21515.33			
Total	8269533	143				

C. Two-way ANOVA of the effects of WS genotype and *wspR* allele on Congo Red binding (Figure 7.6.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Genotype	335954.7	5	67190.94	1.405368	0.239273	2.408513
Allele	460374.1	3	153458	3.209734	0.031141	2.79806
Interaction	475324.1	15	31688.27	0.662793	0.80646	1.880174
Within	2294889	48	47810.19			
Total	3566542	71				

D. One-way ANOVA of the effect of WS genotype on Congo Red binding (Figure 7.6.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Genotypes	72131.04	9	8014.56	4.465184	0.002569	2.392817
Within Genotypes	35898.01	20	1794.9			
Total	108029	29				

E. Two-way ANOVA of the effects of genotype and *wspR* allele on fitness (Figure 7.7.):

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Allele	1.186114	3	0.395371	7.081196	0.000492	2.79806
Genotype	0.769945	5	0.153989	2.757979	0.028677	2.408513
Interaction	2.17177	15	0.144785	2.593127	0.006326	1.880174
Within	2.680032	48	0.055834			
Total	6.807861	71				