

1 Sociomics: using omic approaches to understand social evolution

2

3 Melanie Ghoul^{1*}, Sandra B. Andersen^{2*} and Stuart A. West¹

4 Melanie.ghoul@zoo.ox.ac.uk

5 sandrabreumandersen@gmail.com

6 stuart.west@zoo.ox.ac.uk

7 *Authors have contributed equally to this work

8 ¹Zoology Department of zoology, University of Oxford, Oxford, OX1 3PS

9 ²Langone Medical Center, New York University, 423 East 23rd Street, New York, NY 10010

10

11

12 Keywords: altruism, cooperation, conflict, genomics, transcriptomics, proteomics, metabolomics

13

14

15 Abstract

16

17 All of life is social, from genes cooperating to form organisms, to animals cooperating to form
18 societies. Omic approaches offer exceptional opportunities to solve major outstanding problems in
19 the study of how sociality evolves. First, omics can be used to clarify the extent and form of
20 sociality in natural populations. This is especially useful in species where it is hard to study social
21 traits in natural populations, such as bacteria and other microbes. Second, omics can be used to
22 examine the consequences of sociality for genome evolution and gene expression. This is especially
23 useful in cases where there is clear variation in the level of sociality, such as the social insects.
24 Major tasks for the future are to apply these approaches to a wider range of non-model organisms,
25 and to move from exploratory analyses to the testing of evolutionary theory.

26

27 Bringing social evolution studies into the omics era

28

29 Social traits have fitness consequences for both the individual that performs the behavior and
30 another individual [1]. From an evolutionary perspective, social traits pose a number of problems.
31 For example, why would individuals perform cooperative behaviours, which benefit other
32 individuals, and hence could decrease the relative fitness of the cooperator? A large body of
33 theoretical and empirical work has addressed this issue, showing that explanations for **cooperation**
34 can be divided into two categories [2-4]. First, altruistic cooperation can be favoured if it is directed
35 towards relatives, who share genes, and hence provides kin selected or indirect benefits [1]. Second,
36 cooperation can be favoured towards non-relatives, even members of others species, if it feeds back

37 some direct benefit to the cooperator [5]. For example, if cooperators are rewarded with
38 cooperation, or if non-cooperators are punished. The empirical research in this area has primarily
39 been at the behavioural or phenotype level, often using cost-benefit analyses to examine why traits
40 such as cooperation have been favoured, across a range of organisms, including bacteria, insects,
41 birds and mammals [2-4, 6].

42 In contrast, research on social traits has made relatively little use of “**omic**”
43 methodologies. Omic approaches include high throughput sequencing for (meta)- **genomic**
44 analyses, deep sequencing of targeted genes for diversity analyses, identification of **methylation**
45 epigenetic markers, **transcriptomics**, for inference of gene expression, **metabolomics and**
46 **proteomics** for measuring productivity. These techniques have revolutionized the way we view
47 organisms. Omic studies allow us to detect variation and changes from the level of nucleic acids to
48 the proteins and metabolites produced and link them to the environment. For example, genomic
49 studies have found the genes associated with virulence and antibiotic resistance in pathogens, and
50 transcriptomics has shown how the expression and regulation of these genes depends on whether
51 the pathogen is inside or outside of its host [7, 8]. These methodologies have not been used widely
52 to provide the kind of estimates of costs and benefits that are the focus of social evolution research.

53 Omic approaches, however, offer a number of unique opportunities for resolving
54 outstanding problems in our understanding of how and why social traits evolve. We suggest two
55 broad questions, which can be answered with omic methodologies. First, what is the extent and
56 form of sociality in natural populations? In some organisms, such as insects or birds, the nature of
57 social interactions is relatively obvious [2, 3, 9]. In contrast, in other organisms, such as bacteria
58 living as **symbionts**, or slime moulds living in the soil, it can be hard to know what social
59 interactions are really going on. Second, what are the omic consequences of sociality? This involves
60 a variety of issues from what kind of genetic changes are associated with transitions to more social
61 life [10, 11], to how gene expression can lead to individuals with the same genotype having very
62 different phenotypes (**division of labour**) [12-14]. We highlight the state of the art of the field,
63 commonalities across systems, and how insights can be applied across taxa. Our examples are
64 biased towards microbes and social insects, where these approaches have been most used.

65 66 **What is the extent and form of sociality in natural populations?**

67
68 Before we can start to think about the costs and benefits of cooperation, we need to know who is
69 interacting, and what they are doing for each other. Sequencing can be used to discover and identify
70 microsymbionts [15]. Whole genome sequencing, transcriptomics, metabolomics, and proteomics
71 can be used to determine which metabolic pathways are used and therefore how individuals are

72 interacting [16-18].

73

74 *Who is interacting?*

75 Sequencing facilitates the discovery of new symbionts [15]. Lichens are a symbiosis between fungi
76 and algae and/or cyanobacteria. Lichens vary in the production of vulpinic acid, which is a toxin
77 that likely serves as a herbivore repellent, and cause a distinct color change of the lichen from
78 brown to bright yellow. However, genomic and transcriptomic analyses of sequences identified as
79 coming from the fungal and photosynthetic partners, failed to identify any genetic variation
80 correlated with the production of vulpinic acid. By examining the remaining sequences, that would
81 otherwise be removed as “noise”, it was found that a third partner was also present in the mutualism
82 – a basidiomycete yeast. The abundance of this yeast correlates with the production of vulpinic
83 acid, and thus host phenotype. Furthermore, this yeast seems to be a common partner in the
84 mutualism, with additional screening showing related yeasts in 52 genera of lichens from six
85 continents. These yeasts form a monophyletic clade with the closest known relative being a lichen
86 parasite. Sequencing has, therefore, forced us to reevaluate our understanding of an iconic
87 symbiosis, which had been intensely studied for more than 140 years [15] (Figure 1).

88

89 *How are they helping?*

90 Whole genome analyses can be used to investigate the metabolic functions of symbionts. An
91 example is provided by work on the bacterial symbionts associated with a sap-feeding sharpshooter,
92 *Homalodisca vitripennis* [17, 19]. Sap is a poor source of nutrients and cannot sustain the
93 sharpshooter alone, but this insect contains two bacteria, from the genera *Baumannia* and *Sulcia*.
94 The genomes of the symbionts were assembled with a metagenomic approach, and analyses
95 suggested that these symbionts provide complementary functions, with *Sulcia* synthesizing essential
96 amino acids, and *Baumannia* synthesizing various vitamins and cofactors necessary for the host.
97 There is also a striking metabolic complementarity between these bacteria. For example, *Sulcia*
98 synthesizes homoserine, which is then used by *Baumannia* to synthesise the essential amino acid
99 methionine for the host. Similarly, *Baumannia* provides the polyisoprenoids, which *Sulcia* then uses
100 to synthesise menaquinone (Figure1, Key Figure). In the *Hodgkinia* endosymbionts of cicadas,
101 single lineages have diverged into multiple complementary symbionts, which together provide their
102 hosts with the services that the original symbiont contributed alone [20, 21].

103 Proteomics and metabolomics can be used to identify which pathways are active in an
104 association. For example, *Olavius algarvensis* is a worm that lacks both a mouth and a gut, and
105 lives in the sediment at the bottom of the sea, moving between layers where oxygen is present or
106 limited. This worm contains four different symbionts [16]. Metaproteomics and metabolomics

107 showed that three of the symbionts are capable of using carbon monoxide as an energy source,
108 which had not previously been observed in chemosynthetic symbionts. When residing in the anoxic
109 part of the sediment the host metabolism is anaerobic and produces fermentative waste products
110 such as acetate and propionate. It was found that one symbiont recycled these into carbon storage in
111 the form of glycogen or polyhydroxyalkanoate. The symbiont appears to have acquired this
112 pathway in relation to the symbiosis, as it is absent in a free-living relative, and its presence was
113 previously overlooked when only the genomic sequences were considered [16]. The human gut
114 microbiome is another area where omic analyses are revealing who is interacting and what they are
115 doing [22].

116

117 *Who is in control?*

118 Proteomics can be used to identify mechanisms by which hosts and symbionts can manipulate each
119 other [23, 24]. For example, the trypanosomatid *Angomonas deanei* is a unicellular parasite of
120 insects and harbours a β -proteobacterial endosymbiont. The trypanosomatid relies on its
121 endosymbiont for synthesis of key metabolites. Proteomic analyses have shown that the
122 trypanosomatid trafficks a specific protein (ETP1) into its endosymbiont's intracellular space [25].
123 This indicates that the host is using sorting machinery to deliver the proteins to its endosymbiont,
124 potentially mediating or synchronising the symbiont's growth cycle.

125

126 *Social viruses?*

127 Viruses represent another group of organisms where omics can provide insights into the basic
128 nature of social interactions. It has typically been assumed that when viruses spread via particles
129 (virions), that there is only one virus genotype per particle, limiting the potential for social
130 interactions. Recent studies, however, have started to overturn this idea. For example, a
131 combination of single-cell isolation and ultra-deep sequencing revealed that, in the vesicular
132 stomatitis virus (VSV), particles can contain multiple genomes [26]. In this case, sequencing has the
133 potential to completely change our view of how social an organism is [27].

134

135 **What are the social dynamics in natural populations?**

136

137 Sociogenomics allows us to explore the dynamics and evolutionary history of social interactions in
138 natural populations [28-30]. The problem of explaining cooperation is commonly framed in terms
139 of understanding the selective forces that prevent the invasion of 'cheats', who do not perform the
140 cooperative behavior, but are able to exploit the cooperative behavior of others [31]. An unresolved
141 issue is the extent to which **cheating** occurs in natural populations, and if so, what are the

142 evolutionary dynamics of cheats [31, 32]. Omics can identify cheats, and infer evolutionary
143 dynamics.

144

145 *Cheating slime moulds*

146 Amoebozoa range from obligate unicellular to conditionally social. The socially advanced
147 species, such as the cellular slime mould *Dictyostelium discoideum*, can form mobile multicellular
148 slugs and fruiting bodies in adverse conditions [33]. The fruiting body is made up of sterile stalk
149 cells, which hold aloft the fertile spore cells. Genotypes that produce a higher proportion of spore
150 cells, and which could therefore represent cheats, are found in natural populations [34]. Are these
151 cheats evolutionarily successful?

152 To look for signatures of selection for cheating, one study examined 150 loci that laboratory
153 experiments had suggested could be involved in cheating, and compared their variation with other
154 areas of the genome [28]. They suggested four evolutionary scenarios (Figure 2), which made
155 contrasting predictions on within- and between species diversity for genes potentially involved in
156 cheating:

- 157 1) Cheats could select for resistance, which would then select for greater cheating, with
158 an escalating arms race. This would lead to repeated selective sweeps of
159 cheating/resistance alleles, which would reduce variation within species and increase
160 divergence between species (Figure 2A).
- 161 2) Cheaters could have a selective advantage when rare, because as they increase in
162 frequency there are fewer cooperators to exploit [35]. In this case, **negative-**
163 **frequency dependence** would maintain both cheats and cooperators in the population.
164 This **balancing selection** would increase variation within species, whilst decreasing
165 divergence between species (Figure 2B).
- 166 3) If relatedness (r) is high in the fruiting body, then conflict is reduced, and so we
167 expect natural selection to maintain a certain level of cooperation. In this case,
168 mutants would be rapidly selected against when they arise (**purifying selection**),
169 leading to reduced variation both within and between species (Figure 2C). This
170 scenario can be thought of as a null hypothesis for selection at functional loci – cheats
171 would be rapidly selected against, and rare.
- 172 4) If the multicellular slugs rarely occur in nature then there would only be weak
173 selection for or against cheating. In this case, with relaxed selection, we would expect
174 increased variation both within and between species. This scenario can be thought of
175 as a null hypothesis for non-functional loci.

176 In the study population, relatively high levels of within-species diversity were found, with
177 decreased between-species diversity for the “cheating genes” [28]. This pattern supports the second
178 hypothesis, that cheats have an advantage when rare, and are maintained in the population by
179 balancing selection. A role of balancing selection was supported by a number of other results,
180 including high rates of non-synonymous variation, low number of haplotypes, and low F_{ST} values.
181 These analyses provide an elegant example of how to decipher social dynamics in a natural microbe
182 population (Box 1).

183

184 *Scavenging for iron*

185 Another area where genomics has been used to study social dynamics is the production of iron
186 scavenging molecules, termed siderophores (Figure 1). Iron availability can limit the growth of
187 bacteria, and in response to this, many species of bacteria excrete siderophores. Laboratory
188 experiments have shown that siderophore production can be cooperative, with the benefits being
189 shared amongst the local population of cells [36]. However, the extent to which siderophore
190 production is cooperative in natural populations, and can be exploited by cheats that do not produce
191 siderophores, was relatively unclear.

192 The genetic basis of siderophore production was analysed in natural isolates of marine
193 bacteria of the *Vibrio* species [37]. It was found that strains that had lost the genes to produce
194 siderophores maintained the receptor for uptake, indicating selection to be able to exploit the
195 siderophores produced by other strains (cheat). Whole genome sequencing also showed that
196 selection to cheat influences iron uptake in the opportunistic pathogen *Pseudomonas aeruginosa*,
197 causing chronic lung infections of cystic fibrosis patients (Figure 3). Using longitudinally sampled
198 isolates from patients it was shown that mutational patterns were consistent with social interactions
199 driving selection on the siderophore genes [38]. Loss of production was most frequently achieved
200 by knock-out mutations of a small gene effectively abolishing production of the siderophore. The
201 receptor for siderophore uptake was, however, maintained in non-producers, only when in the
202 presence of producers they could cheat on (Figure 3).

203

204 **What are the omic consequences of sociality?**

205

206 Transitions to social life can involve numerous changes at the phenotypic level. Compare a solitary
207 wasp with a colony of ants, or free-living bacteria with a complex multicellular organism. How are
208 these changes created at the omic level? Is social change driven by novel gene acquisition or re-
209 wiring of existing expression networks? What are the roles of gene duplication, inversion, mutation,
210 **horizontal gene transfer** or **methylation**?

211

212 *Gene expansion*

213 Social insects use pheromones for complex chemical communication. Whole genome comparisons
214 have shown that the genes for odorant receptors have expanded drastically in the ancestor of ants,
215 followed by further species-specific duplications [39]. The importance of odorant receptors for
216 social life was investigated in the clonal raider ant *Ooceraea biroi*. Using CRISPR, a knock-out was
217 created of the gene for the co-receptor protein Orco, which in complex with odorant receptors
218 compose the functional sensing unit. The knock-out ants exhibited reduced capabilities in
219 responding to odours, in the form of repellent and pheromone trails used for foraging, and
220 decreased contact with nest-mates. Overall fitness was reduced with a shorter life-span and lower
221 reproductive output (W. Tribble et al., unpublished). This *orco* mutant is the first genetically
222 modified ant, and represents an exciting new approach that could be applied to test numerous other
223 hypotheses.

224

225 *Rewiring of ancestral networks*

226 Comparative studies across a range of taxa suggest that the evolution of gene regulation, rather than
227 acquisition of novel genes or gene functions, is key to the evolution of sociality [10, 11, 40]. Across
228 ten bee species, the level of sociality was compared with that of regulatory complexity [11]. It was
229 found that more social species had: (i) more binding sites genome-wide for **transcription factors**,
230 the proteins that bind to DNA to regulate gene expression; (ii) a higher predicted fraction of
231 methylated genes, where the addition of a methyl group to the DNA modifies gene function and
232 controls gene expression. Further, genes involved in gene regulation, e.g. in transcription,
233 translation and RNA splicing, evolved faster in more social species. This suggests that the more
234 social bees have increased regulatory potential, with genes arranged in more complex networks,
235 potentially allowing for greater flexibility to respond to social interactions [11]. Similar patterns
236 have been found for ants [41].

237 The genomes of Amoebozoa were compared, across the range from obligate
238 unicellular to species that form cooperative fruiting bodies [10]. Using whole genome sequencing,
239 and transcriptomics, it was found that the majority of the genes involved in the development of
240 fruiting bodies were conserved and also found in unicellular species. For example, genes
241 responsible for surface adhesion in unicellular species were found to be involved in cell-to-cell
242 adhesion in the social species [10]. Only about 24% of the genes involved in fruiting body
243 formation were unique to social species, and the majority of these were involved in external cell-to-
244 cell communication and recognition. For the volvocine algae, novel genes also only play a limited
245 role in the evolution of multicellularity [40] (Figure 1). Similarly, the evolution of increased

246 sociality in great tits, represented by increased capacity for social learning, appears to be driven by
247 selection on existing genes involved in cognition, rather than new genes [42].

248 Bacteria offer excellent opportunities for examining the omics of social change.

249 Whole-genome sequencing has demonstrated that the transition to a symbiont lifestyle is associated
250 with genome reduction [43]. Furthermore, this pattern has been linked with a cost-benefit approach,
251 to show that symbionts with smaller genomes provide greater benefits to their hosts (R. Fisher et al.,
252 unpublished). In contrast, we know relatively little about the consequences of variation in the level
253 of sociality at the intraspecies level. Does variation in the number and type of regulatory elements
254 reflect variation in sociality [44-46]? Questions to address include how symbiont genome reduction
255 affects intraspecies interactions – is competition between symbionts lowered to the host’s benefit?
256 Also, does the mode of regulation affect sociality? Numerous obligate symbionts control protein
257 production through phase-variable gene expression. Here, genes are turned on or off at random by
258 replication errors caused by slipped-strand mispairing in repetitive regions [47]. Whether this
259 affects social interactions remains to be explored.

260

261 *Gene inversions and supergenes*

262 Research on the fire ant, *Solenopsis invicta*, has suggested that a “supergene” drives the form of
263 sociality. A supergene is defined as a set of genes on a chromosome that are closely linked and
264 therefore inherited together. The fire ant can form colonies with either a single (monogynous) or
265 multiple (polygynous) queens [48]. This difference is driven by an odorant binding protein coded
266 by the gene, *Gp-9*, located on the social B/b chromosomes, where expression differences in many
267 genes leads to divergent phenotypes. A supergene is possible in this system because a genomic
268 rearrangement caused by an inversion prevents recombination in that region (Figure 1). Other
269 examples of supergenes influencing social behavior have been found, including determination of
270 the colony queen number in the alpine silver ant, *Formica selysi* [49], and the male mating behavior
271 in a bird, the ruff, *Philomachus pugnax* [50]. Genomics has allowed us to discover that gene
272 inversions, and the subsequent lack of recombination in those regions maintains social forms that
273 would otherwise not exist.

274

275 *Horizontal gene transfer*

276 In bacteria, there is the potential for significant horizontal gene transfer via conjugation and
277 transformation. This can affect sociality in at least two ways. First, the evolution of sociality can
278 occur by horizontal gene transfer and not through mutations. For example, the genes required for
279 free living rhizobia to transition to become root symbionts occurs through horizontal transfer of
280 genomic islands that carry the symbiosis genes [51, 52]. Second, evolutionary theory suggests we

might be more likely to observe social traits on horizontally transferred genes [53, 54]. The reason for this is that these genes could ‘reinfect’ cheats with cooperative traits, in a way that helps maintain social behaviours. It has been shown that in 21 *E. coli* genomes, the genes encoding secreted proteins, which are more likely to represent social traits, are more likely to be found on mobile elements [55]. However, alternative explanations are possible. For example, we might expect genes involved with adaptation to novel environments to be more likely to be horizontally acquired, and secreted proteins might be especially important in adaptation to new environments.

Acquisition of novel genes can also occur through social interactions between bacteria. For example, *Vibrio cholerae* use a type VI secretion “stabbing” system, to inject toxins into their competitors. This killing mechanism was revealed to be co-regulated with genes required for horizontal gene transfer, therefore allowing the killer cells to also acquire genes from those they stab [56].

Division of labour

A key feature of many societies is the division of labour, when cooperating individuals specialize to carry out specific tasks [14]. This can be observed at all levels of biological organisation, from cells in microbial communities, to cells within multicellular organisms, and multicellular organisms within social insect colonies. How are these different phenotypes produced within societies where the different individuals are either genetically identical, or at least relatively genetically homogenous?

Omics research on bacteria has shown how variation within a clonal population can be caused by random fluctuations in biochemical reactions being amplified by gene networks (bistability or phenotypic noise) [14]. For example, in the intestinal pathogen *Salmonella enterica* serovar Typhimurium, genomic and transcriptomic tools revealed that bistability drives the division of labour between cells that remain in the host gut lumen to reproduce, and those that invade the host tissue to trigger an immune response that eliminates competing bacteria [57] (Figure 1). In other species, signaling between cells is used to divide labour. For example, in cyanobacteria, such as *Anabaena* species, genomics and transcriptomics showed that signaling peptides are transported between cells [58]. This serves to coordinate which cells will contribute to photosynthesis and which will develop into nitrogen fixing heterocysts [58]. The discovery of these various types of mechanisms to generate phenotypic heterogeneity, especially phenotypic noise versus signaling, raises the question of why evolution would favour different mechanisms in different systems [14].

Several studies on social insects have looked for genes that are differentially expressed between worker and reproductive castes. A primary aim has been to resolve if independent origins of sociality have followed the same route to caste differentiation, so that

316 conserved genes or pathways are used, or whether novel taxon-specific genes are more important
317 [13, 59]. A comparative study was done on transcriptomes from the fire ant *Solenopsis invicta*, the
318 honey bee *Apis mellifera* and the paper wasp *Polistes metricus* [12]. Little overlap was found across
319 species in what genes were differentially expressed between castes. Even when there was overlap,
320 expression was not necessarily changed in the same direction. By grouping genes into functional
321 groups, however, the study found that even though the specific genes did not overlap, there were
322 specific pathways and molecular functions that were targeted in all three species, such as the
323 glycolysis/glycogenesis metabolism (Figure 1). There was therefore some evidence of convergent
324 evolution at the functional level.

325 Recent theoretical work proposed a shift from testing developmental hypotheses, such
326 as whether evolution is convergent between species, to more explicitly testing social evolution
327 theory [60, 61]. As genes vary in the extent to which they have direct or indirect (through relatives)
328 fitness effects, this is expected to leave different signatures of selection. For example, in a social
329 insect colony with a singly-mated queen, a gene expressed in sterile workers that increases fitness
330 in its reproductive sisters is likely to experience stronger selection, due to their high relatedness,
331 than one that affects its mother or brother. In contrast, in a colony with a multiply-mated queen, the
332 gene affecting sisters experience the same low selection as one that affects brothers, due to lower
333 relatedness (Figure 4). In practice, this would require (i) identification of genes that are
334 differentially expressed in specific castes, in specific interactions; and (ii) testing whether genetic
335 variation in these genes, compared to others, show patterns of positive selection. A start has been
336 made to use this approach across different systems, with mixed results. In the pharaoh ant
337 *Monomorium pharaonis*, with multiple queens, genes upregulated in workers were found to
338 experience reduced selection compared to genes upregulated in queens (M. Warner et al., In Press).
339 The opposite was found for worker genes in the honey bee [62, 63]. This may highlight the
340 difficulties in distinguishing between relaxed purifying selection and positive selection, or reflect
341 life history differences between the systems.

342

343 *Concluding Remarks*

344 We have only just begun to exploit sociomics. Omic technology has been able to clarify
345 fundamental aspects of socio-biology including who is interacting, and how they are interacting.
346 However, sociomic studies have often used an exploratory approach to identify variations among
347 individuals' omic profiles to then speculatively explain differences in social behaviours. We are
348 now at the stage where experiments can be designed to explicitly link the occurrence of social
349 behaviours with variation across the genome, transcriptome and metabolome (Box 1 and
350 Outstanding Questions). Are changes observed at the genome level actually influencing the

351 phenotypic transcriptome and/or metabolome? Do we see patterns of omic signatures across taxa
352 (phylogenies) for similar behaviours? Subsequently how does this influence the costs and benefits
353 of the social interactions studied?

354

355 Glossary

356

357 **Balancing selection:** when multiple alleles are maintained in the population conserving genetic
358 polymorphism, and is usually due to frequency-dependent selection or heterozygote advantage.

359

360 **Cheating:** a trait that is beneficial to a cheat and costly to a cooperator in terms of inclusive fitness,
361 when these benefits and costs arise from the actor directing a cooperative behavior toward the
362 cheat, rather than the intended recipient.

363 **Cooperation:** a behaviour is cooperative if it provides a benefit to another individual and if it has
364 evolved at least partially because of this benefit.

365 **Division of labour:** cooperating individuals specialising to carry out specific tasks. Division may
366 occur within a group of organisms or between cells.

367 **Genomics:** the analysis of whole genomes, by DNA sequencing and *denovo* assembly or
368 comparison to a reference genome. This may serve to identify variation between individuals e.g.
369 identify gene loss, discover new genes, mutations, and gene order.

370

371 **Horizontal gene transfer:** the movement of genetic material between organisms that are not parent
372 and offspring. This occurs frequently in bacteria through the uptake of plasmids or genomic DNA.

373

374 **Metabolomics:** the analysis of produced small-molecule metabolites (<1500 Da) present e.g. in
375 a specific tissue or organism.

376

377 **Meta-omics:** High-throughput, global analysis of genomic, transcriptomic, proteomic, and/or
378 metabolomics data collected from a community of organisms.

379

380 **Methylation:** an epigenetic mechanism where enzymes add methyl groups to DNA without
381 changes in the sequence, modifying gene function and controlling expression and subsequently
382 modifying the phenotype.

383

384 **Negative-frequency dependent selection:** when the relative fitness of a phenotype or genotype
385 decreases as it becomes more common in a population.

386

387 **Omics approaches:** are large-scale analyses aimed to characterize and quantify what organisms
388 are and how they function, such as transcriptomics, genomics and metabolomics.

389

390 **Proteomics:** the analysis of expressed proteins. This also includes information about translation
391 rates and posttranslational modification of proteins that affect function, such as phosphorylation
392 (addition of phosphate) and ubiquitination (addition of ubiquitin).

393

394 **Purifying selection:** also known as negative selection, which is the removal of deleterious alleles
395 from the population subsequently reducing variation in the population. This can then lead to
396 stabilizing selection.

397

398 **Symbiont:** one of the partners in a symbiosis and the term is typically used to describe the smaller
399 partner.

400
401 **Transcription factors:** proteins that up- or down regulate the expression of genes, by binding to
402 DNA.

403
404 **Transcriptomics:** the analysis of RNA transcripts at the whole genome level, often done by RNA
405 sequencing. This also includes information about transcription rates, gene expression profiles and
406 gene regulation. This is used to analyse which genes are expressed in e.g. specific individuals,
407 tissues or in a given treatment. RNA sequencing may also replace or facilitate *denovo* assembly of
408 whole genomes by focusing on expressed genes instead of large repetitive non-coding regions that
409 may be difficult to assemble.

410
411
412
413

414

415 **Box 1**

416

417 **Challenges with the use of omic tools**

418

419 *Denovo assembly of genomes from mixed samples is problematic*

420

421 Genomics can be used to identify unculturable symbionts or interacting members of multispecies
422 microbial communities. Challenges to *denovo* genome assembly are low sequencing depth of genes
423 from rare species, host tissue contamination in samples from symbioses, and sorting of sequences
424 when multiple symbionts are present. Continued reduction of sequencing costs will allow for deeper
425 sequencing. Partly assembled genomes that have the same frequency in the sample can also be
426 grouped together in analyses - even if these cannot be fully assembled to a closed genome they are
427 likely to come from the same organism [64].

428

429

430 *Within-species variation not considered in between-species comparisons*

431

432 Whole genome comparisons between species are most frequently done with a sample size of one
433 per species, which ignores within-species variation. This may be particularly problematic in
434 microbes where gene content can be highly variable, especially in regions that affect inter- and
435 intraspecies interactions [65, 66].

436

437

438 *Small sample size gives false positives in expression analyses*

439

440 Sample sizes in transcriptomic and methylation analyses are often small, with one sample of pooled
441 individuals, increasing the risk of finding false positives. This problem was highlighted by a study
442 on methylation patterns in clonal raider ant brains, where all individuals in the colony switch
443 between a reproductive- and a worker-like phase. Whilst no significant differences between life
444 stages were found when examining data from four colonies, pair-wise comparisons within colonies
445 would have given positive effects [67].

446

447

448 *Statistical power in comparative studies*

449

450 Comparative genomic analyses, examining the changes that occur with transitions to sociality have
451 often been based on a small number of species, with a limited number of transitions to sociality.
452 Consequentially, in terms of the number of phylogenetically independent transitions to social life,
453 these studies can lack statistical power [68]. The Earth Biogenome Project and BioGenomics2017
454 projects aim to sequence a reference genome from every organism, providing the genomes for
455 numerous species where we know the level of sociality. This will open up numerous avenues for
456 large scale comparative genomic analyses.

457

458

459 *Challenges to the interpretation of omic data*

460

461 To what extent are gene sequences, protein regulation, metabolic data and phenotype correlated?
462 This is particularly challenging in multispecies soil and water microbial communities, where there
463 is a lot of dispersal and nutrient flux affecting community composition over time, and so each
464 sampling event will result in different omic profiles [69, 70]. One solution is to identify key genes
465 that are consistently correlated with traits driving community dynamics, such as the production of
466 public goods molecules, quorum sensing signaling molecules, antagonistic competitive molecules,
467 and metabolic pathways required for substrate use.

468

469 *Generating community structures*

470 Omic data from established communities allows us to interpret what social interactions may have
471 paved the way to generate these communities. Although experiments can then be used to test
472 whether members of the communities can interact in these ways, this does not necessarily mean that

473 they do so in the natural communities. Instead omic tools can be used to track how communities are
474 built over time and what social interactions in fact occurred. For example, a study on marine
475 microbial communities, using transcriptomics, metabolomics and genomics to track the stages of
476 succession and metabolic shifts over time, showed that analyses from a single sampling time had
477 produced misleading results [69].

478

479 *The use of predictive versus exploratory approaches*

480 A general point is that many previous studies have been based on exploratory approaches, where
481 research is focused on finding patterns and correlations. There is need to move to use predictive
482 approaches. Testable hypotheses can be generated from either evolutionary theory directed towards
483 omic approaches [60, 71], or from previous exploratory analyses.

484

485

486

487 **Figure legends**

488

489 Figure 1, Key Figure: An overview of the questions addressed and the tools used. Photo credits:
490 lichen by Tim Wheeler (timwheelerphotography.com); sharpshooter by Daniela Takiya & Roman
491 Rakitov; *Pseudomonas aeruginosa* by Ashleigh S. Griffin; Dictyostelium slime molds by Owen
492 Gilbert; volvocine algae by Aurora Nedelcu; fire ants by S.D. Porter, USDA-ARS; Salmonella from
493 Wikipedia; honey bees by Mark Warren (warrenphotography.co.uk).

494

495 Figure. 2: Evolutionary scenarios predicting cheat-cooperator dynamics in the social amoeba [28].

496 Shaded areas are proportional to the frequencies of different alleles (colors) in a population.

497 A) An escalating arms race results in repeated selective sweeps of cheating or resistance alleles
498 through the population.

499 B) Balancing selection (negative frequency dependence) maintains both cheats and cooperators in a
500 population at stalemate.

501 C) Cheats continually arise through new mutations, but are selected against in a population with
502 high relatedness (purifying selection).

503

504 Figure 3: The social evolution of siderophore production in *Pseudomonas aeruginosa* infections of
505 the cystic fibrosis lung. (i) Early in an infection, bacterial cells (green) produce siderophores
506 (orange) which binds to iron and is taken up as a complex through a specific receptor (blue claw).
507 Genetically, production is initiated by a small sigma-factor (green) that upregulates genes for

production (orange). Expression of the receptor gene (blue) controls uptake. (ii) The population of siderophore producers is invaded by mutants that produce less or no siderophores (grey), but which are still able to uptake the siderophores produced by others as they express the receptor. Mutations for loss of production are biased towards the sigma-factor gene (red cross), which is the most cost-effective way to knock-out the system (iii) Once the cells that produce siderophores (green) have been eliminated from the population, the non-producers (grey) lose the ability to uptake siderophore (claw lost). Mutations accumulate in the receptor gene (red cross).

Figure 4: Predictions for strength of selection on genes with caste specific expression. A gene expressed in a worker ant that affects the fitness of its mother, reproductive sister or brother, respectively, is expected to experience different degrees of selection because of variation in relatedness. In a colony with a singly-mated queen, the worker is most related to their sisters. In contrast, if queens mate multiply, they become relatively less related to their sisters. This is expected to leave different signatures of selection in the specific genes, which may be tested with transcriptomic and whole genome data analyses. Ant drawings by Tim Holtom, antark.net.

Acknowledgements

We would like to thank David Queller, Timothy A. Linksvayer and two anonymous reviewers for their useful comments.

References

1. Hamilton, W.D. (1964) Genetical Evolution of Social Behaviour I & II. *Journal of Theoretical Biology* 7 (1), 1-52.
2. Sachs, J.L. et al. (2004) The evolution of cooperation. *Quarterly Review of Biology* 79 (2), 135-160.
3. West, S.A. et al. (2007) Evolutionary explanations for cooperation. *Current Biology* 17 (16), R661-R672.
4. Bourke, A.F.G. (2011) *Principles of social evolution*, Oxford University Press.
5. Trivers, R.L. (1971) The evolution of reciprocal altruism. *The Quarterly Review of Biology* 46 (1), 35-37.
6. Davies, N.B. et al. (2012) *An introduction to behavioural ecology*, 4th edn., Wiley-Blackwell.
7. Westermann, A.J. et al. (2017) Resolving host-pathogen interactions by dual RNA-seq. *PLoS Pathog* 13 (2), e1006033.
8. Chen, P.E. and Shapiro, B.J. (2015) The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol* 25, 17-24.
9. Bourke, A.F. (2014) Hamilton's rule and the causes of social evolution. *Philos Trans R Soc Lond B Biol Sci* 369 (1642), 20130362.

- 548 10. Glockner, G. et al. (2016) The multicellularity genes of dictyostelid social amoebas. *Nat*
549 *Commun* 7, 12085.
- 550 11. Kapheim, K.M. et al. (2015) Social evolution. Genomic signatures of evolutionary
551 transitions from solitary to group living. *Science* 348 (6239), 1139-43.
- 552 12. Berens, A.J. et al. (2015) Comparative transcriptomics of convergent evolution: different
553 genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects.
554 *Mol Biol Evol* 32 (3), 690-703.
- 555 13. Grozinger, C.M. et al. (2007) Genome-wide analysis reveals differences in brain gene
556 expression patterns associated with caste and reproductive status in honey bees (*Apis*
557 *mellifera*). *Mol Ecol* 16 (22), 4837-48.
- 558 14. West, S.A. and Cooper, G.A. (2016) Division of labour in microorganisms: an evolutionary
559 perspective. *Nat Rev Microbiol* 14 (11), 716-723.
- 560 15. Spribille, T. et al. (2016) Basidiomycete yeasts in the cortex of ascomycete macrolichens.
561 *Science* 353 (6298), 488-92.
- 562 16. Kleiner, M. et al. (2012) Metaproteomics of a gutless marine worm and its symbiotic
563 microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci*
564 *U S A* 109 (19), E1173-82.
- 565 17. McCutcheon, J.P. and Moran, N.A. (2007) Parallel genomic evolution and metabolic
566 interdependence in an ancient symbiosis. *Proc Natl Acad Sci U S A* 104 (49), 19392-7.
- 567 18. Douglas, A.E. (2009) The microbial dimension in insect nutritional ecology. *Functional*
568 *Ecology* 23 (1), 38-47.
- 569 19. Wu, D. et al. (2006) Metabolic complementarity and genomics of the dual bacterial
570 symbiosis of sharpshooters. *PLoS Biol* 4 (6), e188.
- 571 20. Campbell, M.A. et al. (2015) Genome expansion via lineage splitting and genome reduction
572 in the cicada endosymbiont *Hodgkinia*. *Proc Natl Acad Sci U S A* 112 (33), 10192-9.
- 573 21. Van Leuven, J.T. et al. (2014) Sympatric speciation in a bacterial endosymbiont results in
574 two genomes with the functionality of one. *Cell* 158 (6), 1270-80.
- 575 22. Manor, O. et al. (2014) Mapping the inner workings of the microbiome: genomic- and
576 metagenomic-based study of metabolism and metabolic interactions in the human
577 microbiome. *Cell Metab* 20 (5), 742-52.
- 578 23. Login, F.H. et al. (2011) Antimicrobial peptides keep insect endosymbionts under control.
579 *Science* 334 (6054), 362-5.
- 580 24. Van de Velde, W. et al. (2010) Plant peptides govern terminal differentiation of bacteria in
581 symbiosis. *Science* 327 (5969), 1122-6.
- 582 25. Morales, J. et al. (2016) Development of a toolbox to dissect host-endosymbiont
583 interactions and protein trafficking in the trypanosomatid *Angomonas deanei*. *BMC Evol Biol*
584 16 (1), 247.
- 585 26. Combe, M. et al. (2015) Single-Cell Analysis of RNA Virus Infection Identifies Multiple
586 Genetically Diverse Viral Genomes within Single Infectious Units. *Cell Host Microbe* 18 (4),
587 424-32.
- 588 27. Sanjuan, R. (2017) Collective Infectious Units in Viruses. *Trends Microbiol.*
- 589 28. Ostrowski, E.A. et al. (2015) Genomic signatures of cooperation and conflict in the social
590 amoeba. *Curr Biol* 25 (12), 1661-5.
- 591 29. Robinson, G.E. et al. (2005) Sociogenomics: social life in molecular terms. *Nat Rev Genet* 6
592 (4), 257-70.
- 593 30. Foster, K.R. et al. (2007) What can microbial genetics teach sociobiology? *Trends Genet* 23
594 (2), 74-80.
- 595 31. Ghoul, M. et al. (2014) Toward an Evolutionary Definition of Cheating. *Evolution* 68 (2),
596 318-331.
- 597 32. Jones, E.I. et al. (2015) Cheaters must prosper: reconciling theoretical and empirical
598 perspectives on cheating in mutualism. *Ecol Lett.*

599 33. Strassmann, J.E. and Queller, D.C. (2011) Evolution of cooperation and control of cheating
600 in a social microbe. *Proc Natl Acad Sci U S A* 108 Suppl 2, 10855-62.

601 34. Strassmann, J.E. et al. (2000) Altruism and social cheating in the social amoeba
602 *Dictyostelium discoideum*. *Nature* 408 (6815), 965-7.

603 35. Ross-Gillespie, A. et al. (2007) Frequency dependence and cooperation: theory and a test
604 with bacteria. *American Naturalist* 170 (3), 331-342.

605 36. Griffin, A.S. et al. (2004) Cooperation and competition in pathogenic bacteria. *Nature* 430
606 (7003), 1024-1027.

607 37. Cordero, O.X. et al. (2012) Public good dynamics drive evolution of iron acquisition
608 strategies in natural bacterioplankton populations. *Proceedings of the National Academy of*
609 *Sciences of the United States of America* 109 (49), 20059-20064.

610 38. Andersen, S.B. et al. (2015) Long-term social dynamics drive loss of function in pathogenic
611 bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 112
612 (34), 10756-61.

613 39. Zhou, X. et al. (2015) Chemoreceptor Evolution in Hymenoptera and Its Implications for
614 the Evolution of Eusociality. *Genome Biol Evol* 7 (8), 2407-16.

615 40. Hanschen, E.R. et al. (2016) The *Gonium pectorale* genome demonstrates co-option of cell
616 cycle regulation during the evolution of multicellularity. *Nat Commun* 7, 11370.

617 41. Simola, D.F. et al. (2013) Social insect genomes exhibit dramatic evolution in gene
618 composition and regulation while preserving regulatory features linked to sociality. *Genome*
619 *Res* 23 (8), 1235-47.

620 42. Laine, V.N. et al. (2016) Evolutionary signals of selection on cognition from the great tit
621 genome and methylome. *Nat Commun* 7, 10474.

622 43. Moran, N.A. et al. (2008) Genomics and evolution of heritable bacterial symbionts. *Annu*
623 *Rev Genet* 42, 165-90.

624 44. McAdams, H.H. and Arkin, A. (1999) It's a noisy business! Genetic regulation at the
625 nanomolar scale. *Trends Genet* 15 (2), 65-9.

626 45. Perez-Rueda, E. et al. (2004) Phylogenetic distribution of DNA-binding transcription
627 factors in bacteria and archaea. *Comput Biol Chem* 28 (5-6), 341-50.

628 46. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* 136 (4), 615-28.

629 47. Henderson, I.R. et al. (1999) Molecular switches--the ON and OFF of bacterial phase
630 variation. *Mol Microbiol* 33 (5), 919-32.

631 48. Wang, J. et al. (2013) A Y-like social chromosome causes alternative colony organization in
632 fire ants. *Nature* 493 (7434), 664-8.

633 49. Purcell, J. et al. (2014) Convergent genetic architecture underlies social organization in
634 ants. *Curr Biol* 24 (22), 2728-32.

635 50. Kupper, C. et al. (2016) A supergene determines highly divergent male reproductive
636 morphs in the ruff. *Nat Genet* 48 (1), 79-83.

637 51. Haskett, T.L. et al. (2016) Assembly and transfer of tripartite integrative and conjugative
638 genetic elements. *Proc Natl Acad Sci U S A* 113 (43), 12268-12273.

639 52. Sachs, J.L. et al. (2011) Evolutionary transitions in bacterial symbiosis. *Proc Natl Acad Sci*
640 *U S A* 108 Suppl 2, 10800-7.

641 53. Mc Ginty, S.E. et al. (2011) Horizontal gene transfer and the evolution of bacterial
642 cooperation. *Evolution* 65 (1), 21-32.

643 54. Smith, J. (2001) The social evolution of bacterial pathogenesis. *Proc Biol Sci* 268 (1462),
644 61-9.

645 55. Nogueira, T. et al. (2009) Horizontal gene transfer of the secretome drives the evolution of
646 bacterial cooperation and virulence. *Curr Biol* 19 (20), 1683-91.

647 56. Borgeaud, S. et al. (2015) The type VI secretion system of *Vibrio cholerae* fosters
648 horizontal gene transfer. *Science* 347 (6217), 63-7.

649 57. Diard, M. et al. (2013) Stabilization of cooperative virulence by the expression of an
650 avirulent phenotype. *Nature* 494 (7437), 353-356.

651 58. Herrero, A. et al. (2016) The multicellular nature of filamentous heterocyst-forming
652 cyanobacteria. *FEMS Microbiol Rev* 40 (6), 831-854.

653 59. Patalano, S. et al. (2015) Molecular signatures of plastic phenotypes in two eusocial insect
654 species with simple societies. *Proc Natl Acad Sci U S A* 112 (45), 13970-5.

655 60. Linksvayer, T.A., & Wade, M. J. (2016) Theoretical Predictions for Sociogenomic Data: The
656 Effects of Kin Selection and Sex-Limited Expression on the Evolution of Social Insect Genomes.
657 *Frontiers in Ecology and Evolution* 4, 1-10.

658 61. Linksvayer, T.A. and Wade, M.J. (2009) Genes with social effects are expected to harbor
659 more sequence variation within and between species. *Evolution* 63 (7), 1685-96.

660 62. Vojvodic, S. et al. (2015) The transcriptomic and evolutionary signature of social
661 interactions regulating honey bee caste development. *Ecol Evol* 5 (21), 4795-807.

662 63. Harpur, B.A. et al. (2014) Population genomics of the honey bee reveals strong signatures
663 of positive selection on worker traits. *Proc Natl Acad Sci U S A* 111 (7), 2614-9.

664 64. Nielsen, H.B. et al. (2014) Identification and assembly of genomes and genetic elements in
665 complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32 (8), 822-
666 8.

667 65. Smith, E.E. et al. (2005) Evidence for diversifying selection at the pyoverdine locus of
668 *Pseudomonas aeruginosa*. *Journal of Bacteriology* 187 (6), 2138-2147.

669 66. McInerney, J.O. et al. (2017) Why prokaryotes have pangenomes. *Nat Microbiol* 2, 17040.

670 67. Libbrecht, R. et al. (2016) Robust DNA Methylation in the Clonal Raider Ant Brain. *Curr*
671 *Biol* 26 (3), 391-5.

672 68. Fisher, R.M. et al. (2013) Group Formation, Relatedness, and the Evolution of
673 Multicellularity. *Current Biology* 23 (12), 1120-1125.

674 69. Datta, M.S. et al. (2016) Microbial interactions lead to rapid micro-scale successions on
675 model marine particles. *Nat Commun* 7, 11965.

676 70. Myrold, D.D. and Nannipieri, P. (2014) Classical Techniques versus Omics Approaches. In
677 *Omics in Soil Science* (Paolo Nannipieri, G.P.a.G.R. ed), pp. 179-187, Caister Academic Press.

678 71. Gardner, A. (2012) Evolution of maternal care in diploid and haplodiploid populations. *J*
679 *Evol Biol* 25 (8), 1479-86.

680 -Warner, M. R., Mikheyev, A., Linksvayer, T. A. (2016). Genomic signature of kin selection in an
681 ant with obligately sterile workers. *Mol Biol Evo. In press.*

682