Trends in Genetics

CellPress

Review

Sociomics: Using Omic Approaches to Understand Social Evolution

Melanie Ghoul,[1,3,*] Sandra B. Andersen,[2,3,*] and Stuart A. West[1]

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

Trends

Social interactions play a key role at all levels of life, from genes cooperating to form organisms, to males fighting to obtain mates.

Omic technologies provide novel opportunities to resolve outstanding problems for our understanding of how social traits evolve.

Omic tools can be used to clarify the extent and form of sociality in natural populations. These approaches are particularly useful in species where it is difficult to study social traits in natural populations, such as microbes.

Omics can also be used to examine the consequences of sociality for genome evolution and gene expression. This is particularly useful in species where the level of sociality in natural populations is already understood, such as social insects.

Bringing Social Evolution Studies into the Omic Era

Social traits have fitness consequences for both the individual that performs the behavior and for another individual [1]. From an evolutionary perspective, social traits pose several problems. For example, why would individuals perform cooperative behaviors, which benefit other individuals, and hence could decrease the relative fitness of the cooperator? A large body of theoretical and empirical work has addressed this issue, showing that explanations for **cooperation** (see Glossary) can be divided into two categories [2–4]. First, altruistic cooperation can be favored if it is directed towards relatives, who share genes, and hence provides kin selected or indirect benefits [1]. Second, cooperation can be favored towards non-relatives, even members of other species, if it feeds back some direct benefit to the cooperator [5]. For example, if cooperators are rewarded with cooperation, or if non-cooperators are punished. The empirical research in this area has primarily been at the behavioral or phenotypic level, often using cost–benefit analyses to examine why traits such as cooperation have been favored across a range of organisms including bacteria, insects, birds, and mammals [2–4,6].

By contrast, research on social traits has made relatively little use of **omic methodologies**. Omic approaches include high-throughput sequencing for (meta)-**genomic** analyses, deep sequencing of targeted genes for diversity analyses, identification of **methylation** epigenetic markers, **transcriptomics** for inference of gene expression, and **metabolomics and proteomics** for measuring productivity. These techniques have revolutionized the way we view organisms. Omic studies allow us to detect variation and changes from the level of nucleic acids to the proteins and metabolites produced, and link them to the environment. For example, genomic studies have found the genes associated with virulence and antibiotic resistance in pathogens, and transcriptomics has shown how the expression and regulation of these genes

[1]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[2]Langone Medical Center, New York University, 423 East 23rd Street, New York, NY 10010, USA
[3]These authors contributed equally to this work.

*Correspondence:
Melanie.ghoul@zoo.ox.ac.uk (M. Ghoul) and
sandrabreumandersen@gmail.com (S. B. Andersen).

depends on whether the pathogen is inside or outside its host [7,8]. These methodologies have not been used widely to provide the type of estimates of costs and benefits that are the focus of social evolution research.

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

What Is the Extent and Form of Sociality in Natural Populations?

Before we can start to think about the costs and benefits of cooperation, we need to know who is interacting, and what they are doing for each other. Sequencing can be used to discover and identify microsymbionts [15]. Whole-genome sequencing, transcriptomics, metabolomics, and proteomics can be used to determine which metabolic pathways are used and therefore how individuals are interacting [16–18].

Who Is Interacting?

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

How Are They Helping?

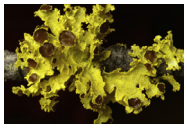
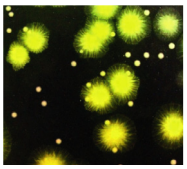
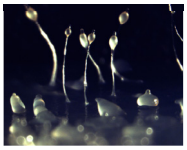
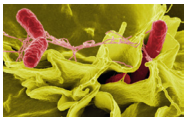
Whole-genome analyses can be used to investigate the metabolic functions of symbionts. An example is provided by work on the bacterial symbionts associated with a sap-feeding sharpshooter, *Homalodisca vitripennis* [17,19]. Sap is a poor source of nutrients and cannot sustain the sharpshooter alone, but this insect contains two bacteria from the genera *Baumannia* and *Sulcia.* The genomes of the symbionts were assembled with a metagenomic approach, and analyses suggested that these symbionts provide complementary functions, with *Sulcia* synthesizing essential amino acids and *Baumannia* synthesizing various vitamins and cofactors necessary for the host. There is also a striking metabolic complementarity between these bacteria. For example, *Sulcia* synthesizes homoserine, which is then used by *Baumannia* to synthesize the essential amino acid methionine for the host. Similarly,

Glossary

Balancing selection: when multiple alleles are maintained in the population, thereby conserving genetic polymorphism; is usually due to frequency-dependent selection or heterozygote advantage.
Cheating: a trait that is beneficial to a cheat and costly to a cooperator, in terms of inclusive fitness, when these benefits and costs arise from the actor directing a cooperative behavior toward the cheat rather than toward the intended recipient.
Cooperation: a behavior is cooperative if it provides a benefit to another individual and if it has evolved at least partially because of this benefit.
Division of labor: cooperation between individuals that are specialized to carry out specific tasks. Division may occur within a group of organisms or between cells.
Genomics: the analysis of whole genomes by DNA sequencing and *de novo* assembly or comparison to a reference genome. This may serve to identify variation between individuals, for example to identify gene loss, discover new genes, mutations, and gene order.
Horizontal gene transfer: the movement of genetic material between organisms that are not parent and offspring. This occurs frequently in bacteria through the uptake of plasmids or genomic DNA.
Metabolomics: the analysis of produced small-molecule metabolites (<1500 Da) present for example in a specific tissue or organism.
Meta-omics: high-throughput, global analysis of genomic, transcriptomic, proteomic, and/or metabolomic data collected from a community of organisms.
Methylation: an epigenetic mechanism where enzymes add methyl groups to DNA without changes in the sequence, modifying gene function and controlling expression and subsequently modifying the phenotype.
Negative frequency-dependent selection: when the relative fitness of a phenotype or genotype decreases as it becomes more common in a population.
Omic methodologies: large-scale analyses used to characterize and quantify what organisms are and how they function, such as

Key Figure

An Overview of the Questions Addressed and the Tools Used

Question	Approach
Who is interacting? Lichens were thought to be a symbiosis between algae and fungus. 	**Transcriptomics** revealed the presence of a third partner, a basidiomycete yeast [15].
What are they doing? Sharpshooters feed on nutrient poor sap. What do its bacterial symbionts supply? 	**Whole-genome sequencing** showed that Sulcia and Baumannia symbionts complement the host and each other in nutrient and vitamin production [17].
Which social dynamics occur? Bacteria and social amoeba can cooperate, by producing siderophores to scavenge iron, and forming multicellular slugs. Do non-cooperative 'cheats' occur and affect dynamics in natural populations?  	**Whole-genome sequencing** showed that cheaters arise in both systems. In the bacteria, cheaters retain the ability to take up siderophores produced by others [37,38]. In the social amoeba, cheaters are maintained in the population by balancing selection and contribute to diversity [28].
What are the omic changes associated with sociality? Volvocine algae range from unicellular to multicellular with division of labor. Colonies of the fire ant *Solenopsis invicta* can accept one or multiple queens. What determines social organization?   	**Whole-genome sequencing** showed that more social algae species had co-opted pathways found in unicellular organisms, e.g., for cell adhesion [40]. **Whole-genome RAD sequencing** of the ants found a gene inversion. Homozygous workers accept one homozygous queen; mixed homo-and heterozygous workers accept multiple heterozygous queens [48].
How is division of labor achieved? How do different phenotypes emerge from genetically similar or identical individuals, such as in *Salmonella enterica* bacteria and honey bees?  	**Whole-genome sequencing and transcriptomics** showed that gene regulation is key. *S. enterica* use bistable expression of virulence genes [57] and for bees, complexity of gene regulation correlates with sociality [11].

Trends in Genetics

(See figure legend on the bottom of the next page.)

transcriptomics, genomics, and metabolomics.
Proteomics: the analysis of expressed proteins. This also includes information about translation rates and post-translational modification of proteins that affect function, such as phosphorylation (addition of phosphate) and ubiquitination (addition of ubiquitin).
Purifying selection: also known as negative selection, the removal of deleterious alleles from the population, subsequently reducing variation in the population. This can then lead to stabilizing selection.
Symbiont: one of the partners in a symbiosis; the term is typically used to describe the smaller partner.
Transcriptomics: the analysis of RNA transcripts at the whole-genome level, often carried out by RNA sequencing. This also includes information about transcription rates, gene expression profiles, and gene regulation. These data are used to analyze which genes are expressed in, for example, specific individuals, tissues, or in a given treatment. RNA sequencing may also replace or facilitate *de novo* assembly of whole genomes by focusing on expressed genes instead of large repetitive non-coding regions that may be difficult to assemble.

Baumannia provides the polyisoprenoids, which *Sulcia* then uses to synthesize menaquinone (Figure 1). In the *Hodgkinia* endosymbionts of cicadas, single lineages have diverged into multiple complementary symbionts which together provide their hosts with the services that the original symbiont contributed alone [20,21].

Proteomics and metabolomics can be used to identify which pathways are active in an association. For example, *Olavius algarvensis* is a worm that lacks both a mouth and a gut, and lives in the sediment at the bottom of the sea, moving between layers where oxygen is present or limited. This worm contains four different symbionts [16]. Metaproteomics and metabolomics showed that three of the symbionts are capable of using carbon monoxide as an energy source, which had not previously been observed in chemosynthetic symbionts. When residing in the anoxic part of the sediment the host metabolism is anaerobic and produces fermentative waste products such as acetate and propionate. It was found that one symbiont recycled these into carbon storage in the form of glycogen or polyhydroxyalkanoate. The symbiont appears to have acquired this pathway in relation to the symbiosis because it is absent in a free-living relative, and its presence was previously overlooked when only the genomic sequences were considered [16]. The human gut microbiome is another area where omic analyses are revealing who is interacting and what they are doing [22].

Who Is in Control?

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

Social Viruses?

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

What Are the Social Dynamics in Natural Populations?

Sociogenomics allows us to explore the dynamics and evolutionary history of social interactions in natural populations [28–30]. The problem of explaining cooperation is commonly framed in terms of understanding the selective forces that prevent the invasion of 'cheats' – who do not perform the cooperative behavior but are able to exploit the cooperative behavior of others [31]. An unresolved issue is the extent to which **cheating** occurs in natural populations and, if it does, what are the evolutionary dynamics of cheats [31,32]. Omics can identify cheats, and infer evolutionary dynamics.

Figure 1. Photo credits: lichen by Tim Wheeler (timwheelerphotography.com); sharpshooter by Daniela Takiya and Roman Rakitov; *Pseudomonas aeruginosa* by Ashleigh S. Griffin; *Dictyostelium* slime molds by Owen Gilbert; volvocine algae by Aurora Nedelcu; fire ants by S.D. Porter, US Department of Agriculture (USDA) Agricultural Research Service; *Salmonella* from Wikipedia; honey bees by Mark Warren (warrenphotography.co.uk) (see [11,15,17,28,37,38,40,48,57])

Cheating Slime Moulds

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

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

(i) Cheats could select for resistance, which would then select for greater cheating, with an escalating arms race. This would lead to repeated selective sweeps of cheating/resistance alleles, which would reduce variation within species and increase divergence between species (Figure 2A).

(ii) Cheats could have a selective advantage when rare because as they increase in frequency there are fewer cooperators to exploit [35]. In this case, **negative frequency-dependent selection** would maintain both cheats and cooperators in the population. This **balancing selection** would increase variation within species while decreasing divergence between species (Figure 2B).

(iii) If relatedness (*r*) is high in the fruiting body, then conflict is reduced, and therefore one would expect natural selection to maintain a certain level of cooperation. In this case, mutants would be rapidly selected against when they arise (**purifying selection**), leading to reduced variation both within and between species (Figure 2C). This scenario can be thought of as a null hypothesis for selection at functional loci – cheats would be rapidly selected against, and rare.

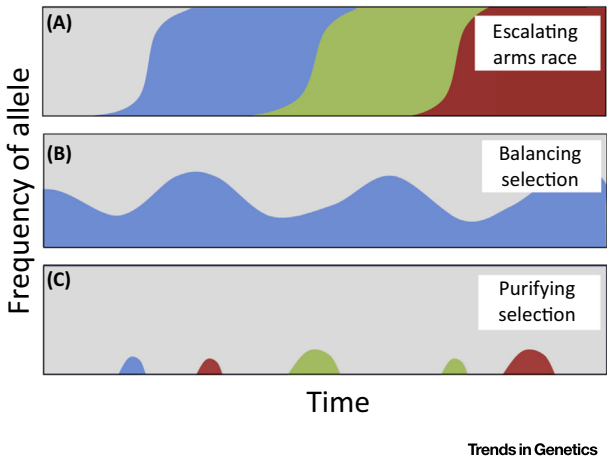


Figure 2. Evolutionary Scenarios Predicting Cheat–Cooperator Dynamics in the Social Amoeba. Figure adapted from [28]. Shaded areas are proportional to the frequencies of the different alleles (colors) in a population. (A) An escalating arms race results in repeated selective sweeps of cheating or resistance alleles through the population. (B) Balancing selection (negative frequency-dependence) maintains both cheats and cooperators in a population at stalemate. (C) Cheats continually arise through new mutations, but are selected against in a population with high relatedness (purifying selection)

(iv) If the multicellular slugs rarely occur in nature then there would only be weak selection for or against cheating. In this case, with relaxed selection we would expect increased variation both within and between species. This scenario can be thought of as a null hypothesis for non-functional loci.

Box 1. Challenges With the Use of Omic Tools

De Novo Assembly of Genomes from Mixed Samples Is Problematic

Genomics can be used to identify unculturable symbionts or interacting members of multispecies microbial communities. Challenges to de novo genome assembly are low sequencing depth of genes from rare species, host tissue contamination in samples from symbioses, and sorting of sequences when multiple symbionts are present. Continued reduction of sequencing costs will allow for deeper sequencing. Partly assembled genomes that have the same frequency in the sample can also be grouped together in analyses - even if these cannot be fully assembled to a closed genome they are likely to come from the same organism [65].

Within-Species Variation Not Considered in Between-Species Comparison

Whole genome comparisons between species are most frequently done with a sample size of one per species, which ignores within-species variation. This may be particularly problematic in microbes where gene content can be highly variable, especially in regions that affect interspecies and intraspecies interactions [66,67].

Small Sample Size Gives False Positives in Expression Analyses

Sample sizes in transcriptomic and methylation analyses are often small, with one sample of pooled individuals, increasing the risk of finding false positives. This problem was highlighted by a study on methylation patterns in clonal raider ant brains, where all individuals in the colony switch between a reproductive- and a worker-like phase. Whilst no significant differences between life stages were found when examining data from four colonies, pair-wise comparisons within colonies would have given positive effects [68].

Statistical Power in Comparative Studies

Comparative genomic analyses examining the changes that occur with transitions to sociality have often been based on a small number of species, with a limited number of transitions to sociality. Consequentially, in terms of the number of phylogenetically independent transitions to social life, these studies can lack statistical power [69]. The Earth Biogenome Project and BioGenomics 2017 projects aim to sequence a reference genome from every organism, providing the genomes for numerous species where we know the level of sociality. This will open up numerous avenues for large-scale comparative genomic analyses.

Challenges to the Interpretation of Omic Data

To what extent are gene sequences, protein regulation, metabolic data, and phenotype correlated? This is particularly challenging in multispecies soil and water microbial communities, where there is a lot of dispersal and nutrient flux affecting community composition over time, and so each sampling event will result in different omic profiles [70,71]. One solution is to identify key genes that are consistently correlated with traits driving community dynamics, such as the production of public goods molecules, quorum sensing signaling molecules, antagonistic competitive molecules, and metabolic pathways required for substrate use.

Generating Community Structures

Omic data from established communities allows us to interpret what social interactions may have paved the way to generate these communities. Although experiments can then be used to test whether members of the communities can interact in these ways, this does not necessarily mean that they do so in the natural communities. Instead, omic tools can be used to track how communities are built over time and what social interactions in fact occurred. For example, a study on marine microbial communities, using transcriptomics, metabolomics, and genomics to track the stages of succession and metabolic shifts over time, showed that analyses from a single sampling time had produced misleading results [70].

The Use of Predictive versus Exploratory Approaches

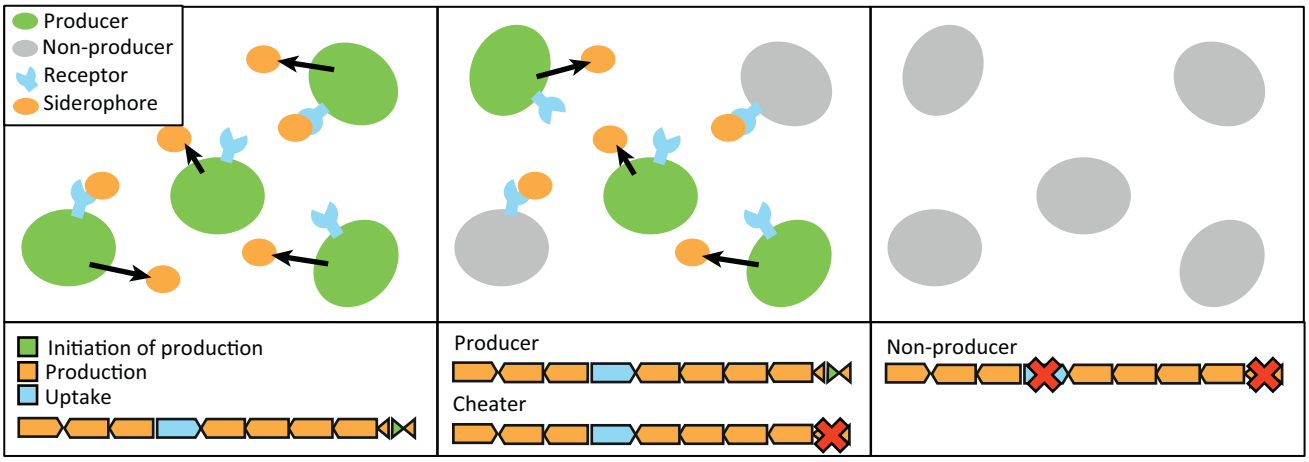
A general point is that many previous studies have been based on exploratory approaches, where research is focused on finding patterns and correlations. There is a need to move to use predictive approaches. Testable hypotheses can be generated from either evolutionary theory directed towards omic approaches [60,72], or from previous exploratory analyses.

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

Scavenging for Iron

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

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



Trends in Genetics

Figure 3. The Social Evolution of Siderophore Production in *Pseudomonas aeruginosa* Infections of the Cystic Fibrosis Lung. (i) Early in an infection, bacterial cells (green) produce siderophores (orange) which bind to iron and are taken up as a complex through a specific receptor (blue claw). 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 because 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 knockout 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), and mutations accumulate in the receptor gene (red cross)

What Are the Omic Consequences of Sociality?

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

Gene Expansion

Social insects use pheromones for complex chemical communication. Whole-genome comparisons have shown that the genes for odorant receptors have expanded drastically in the ancestor of ants, followed by further species-specific duplications [39]. The importance of odorant receptors for social life was investigated in the clonal raider ant *Ooceraea biroi*. Using CRISPR, a knockout was created of the gene for the coreceptor protein Orco which, in complex with odorant receptors, constitutes the functional sensing unit. The knockout ants exhibited reduced abilities to respond to odors, in the form of repellent and pheromone trails used for foraging, and decreased contact with nestmates. Overall fitness was reduced with a shorter lifespan and lower reproductive output (W. Trible *et al.*, unpublished). This *orco* mutant is the first genetically modified ant, and represents an exciting new approach that could be applied to test numerous other hypotheses.

Rewiring of Ancestral Networks

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

The genomes of Amoebozoa were compared across the range from obligate unicellular to species that form cooperative fruiting bodies [10]. Using whole-genome sequencing and transcriptomics, it was found that the majority of the genes involved in the development of fruiting bodies are conserved and are found in unicellular species. For example, genes responsible for surface adhesion in unicellular species were found to be involved in cell-to-cell adhesion in the social species [10]. Only ~24% of the genes involved in fruiting body formation were unique to social species, and the majority of these were involved in external cell-to-cell communication and recognition. For the volvocine algae, novel genes also only play a limited role in the evolution of multicellularity [40] (Figure 1). Similarly, the evolution of increased sociality in great tits, represented by increased capacity for social learning, appears to be driven by selection on existing genes involved in cognition rather than on new genes [42].

Bacteria offer excellent opportunities for examining the omics of social change. Whole-genome sequencing has demonstrated that the transition to a symbiont lifestyle is associated with genome reduction [43]. Furthermore, this pattern has been linked to a cost–benefit approach to show that symbionts with smaller genomes provide greater benefits to their hosts (R. Fisher *et al.*, unpublished). By contrast, we know relatively little about the consequences of variation in the level of sociality at the intraspecies level. Does variation in the number and type of regulatory elements reflect variation in sociality [44–46]? Questions to address include how symbiont

genome reduction affects intraspecies interactions – is competition between symbionts lowered to the benefit of the host? In addition, does the mode of regulation affect sociality? Numerous obligate symbionts control protein production through phase-variable gene expression. In this case genes are turned on or off at random by replication errors caused by slipped-strand mispairing in repetitive regions [47]. Whether this affects social interactions remains to be explored.

Gene Inversions and Supergenes

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

Horizontal Gene Transfer

In bacteria, there is potential for significant horizontal gene transfer via conjugation and transformation. This can affect sociality in at least two ways. First, the evolution of sociality can occur by horizontal gene transfer and not through mutations. For example, the genes required for free-living rhizobia to transition to become root symbionts occurs through horizontal transfer of genomic islands that carry the symbiosis genes [51,52]. Second, evolutionary theory suggests that social traits are more likely to be associated with 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 behaviors. 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, one might expect genes involved with adaptation to novel environments to more likely 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* cells use a type VI secretion 'stabbing' system to inject toxins into their competitors. This killing mechanism was revealed to be coregulated with genes required for horizontal gene transfer, therefore allowing the killer cells to also acquire genes from those they stab [56].

Division of Labor

A key feature of many societies is the division of labor, where cooperating individuals specialize to carry out specific tasks [14]. This can be observed at all levels of biological organization 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?

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

Recent theoretical work proposed a shift from testing developmental hypotheses, such as whether evolution is convergent between species, to more explicitly testing social evolution theory [60,61]. Because genes vary in the extent to which they have direct or indirect (through relatives) fitness effects, this is expected to leave different signatures of selection. For example,

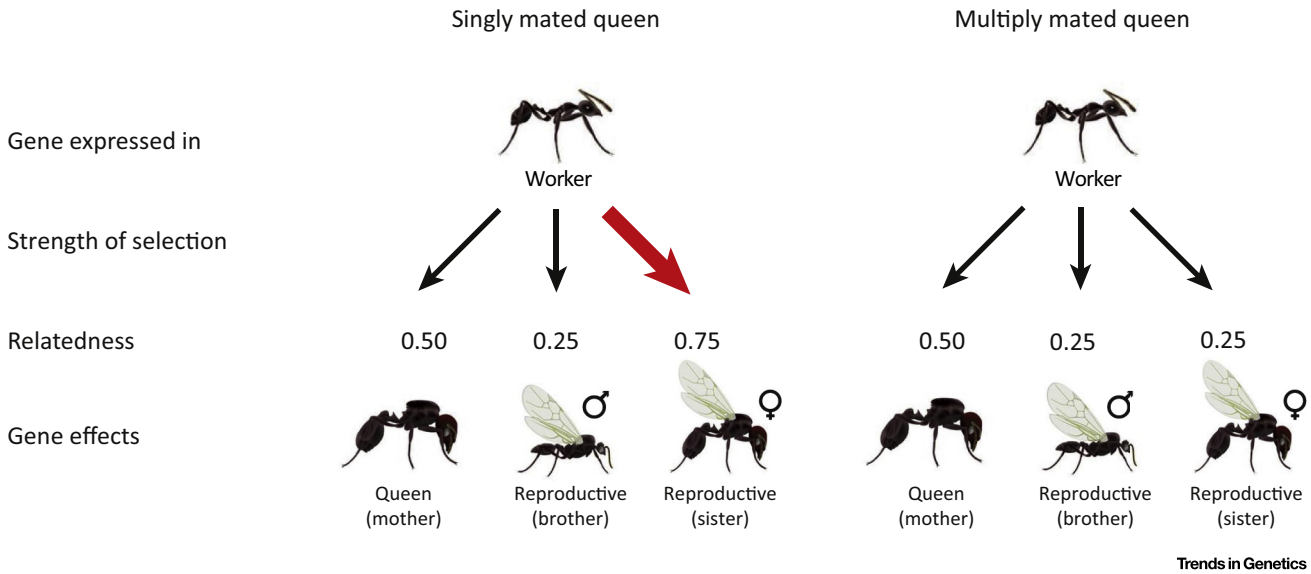


Figure 4. Predictions for Strength of Selection on Genes with Caste-Specific Expression. A gene expressed in a worker ant that affects the fitness of its mother, reproductive sister or brother, respectively, is expected to experience different degrees of selection because of variation in relatedness.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

in a social insect colony with a singly mated queen, a gene expressed in sterile workers that increases fitness in its reproductive sisters is likely to experience stronger selection, due to their high relatedness, than one that affects its mother or brother. By contrast, in a colony with a multiply mated queen, the gene affecting sisters experiences the same low selection as a gene that affects brothers because of lower relatedness (Figure 4). In practice, this would require (i) identification of genes that are differentially expressed in specific castes, in specific interactions; and (ii) testing whether genetic variation in these genes, compared to others, shows patterns of positive selection. A start has been made to use this approach across different systems, with mixed results. In the pharaoh ant *Monomorium pharaonis*, with multiple queens, genes upregulated in workers were found to experience reduced selection compared to genes upregulated in queens [62]. The opposite was found for worker genes in the honey bee [63,64]. This may highlight the difficulties in distinguishing between relaxed purifying selection and positive selection, or reflect life-history differences between the systems.

Concluding Remarks

We have only recently begun to exploit sociomics. Omic technology has been able to clarify fundamental aspects of sociobiology including who is interacting and how they are interacting. However, sociomic studies have often used an exploratory approach to identify variations in the omic profiles of individuals to speculatively explain differences in social behaviors. We are now at a stage where experiments can be designed to explicitly link the occurrence of social behaviors with variation across the genome, transcriptome, and metabolome (Box 1 and Outstanding Questions). Do changes observed at the genome level actually influence the phenotypic transcriptome and/or metabolome? Do we see patterns of omic signatures across taxa (phylogenies) for similar behaviors? Subsequently, how does this influence the costs and benefits of the social interactions studied?

Outstanding Questions

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

How common are undiscovered symbionts, and can hosts control or manipulate their symbionts?

To what extent do viruses cooperate?

Do cheats exist in nature?

Is there conflict in natural populations, and how has conflict been resolved?

What are the omic consequences of sociality, and what types of omic change are associated with transitions to different levels of sociality?

How common are social supergenes, and why do supergenes persist?

Does horizontal gene transfer help to maintain cooperative or social traits?

How is division of labor produced? How can we explain variation in the different mechanisms used to produce division of labor?

Do transitions in sociality rely on gene acquisition or on modifications of existing genes?

How important are post-translational modifications for transitions in sociality and division of labor?

What predictions does social evolutionary theory make for omic data?

How can we exploit the rapidly increasing number of species sequenced?

Acknowledgments

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. *J. Theor. Biol.* 7, 1–52
2. Sachs, J.L. *et al.* (2004) The evolution of cooperation. *Q. Rev. Biol.* 79, 135–160
3. West, S.A. *et al.* (2007) Evolutionary explanations for cooperation. *Curr. Biol.* 17, 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. *Q. Rev. Biol.* 46, 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, 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, 20130362
10. Glockner, G. *et al.* (2016) The multicellularity genes of dictyostelid social amoebas. *Nat. Commun.* 7, 12085
11. Kapheim, K.M. *et al.* (2015) Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348, 1139–1143
12. Berens, A.J. *et al.* (2015) Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* 32, 690–703
13. Grozinger, C.M. *et al.* (2007) Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* 16, 4837–4848
14. West, S.A. and Cooper, G.A. (2016) Division of labour in microorganisms: an evolutionary perspective. *Nat. Rev. Microbiol.* 14, 716–723
15. Spribille, T. *et al.* (2016) Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* 353, 488–492
16. Kleiner, M. *et al.* (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1173–E1182
17. McCutcheon, J.P. and Moran, N.A. (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19392–19397
18. Douglas, A.E. (2009) The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23, 38–47
19. Wu, D. *et al.* (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* 4, e188
20. Campbell, M.A. *et al.* (2015) Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont *Hodgkinia*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10192–10199
21. Van Leuven, J.T. *et al.* (2014) Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158, 1270–1280

22. Manor, O. *et al.* (2014) Mapping the inner workings of the microbiome: genomic- and metagenomic-based study of metabolism and metabolic interactions in the human microbiome. *Cell Metab.* 20, 742–752
23. Login, F.H. *et al.* (2011) Antimicrobial peptides keep insect endosymbionts under control. *Science* 334, 362–365
24. Van de Velde, W. *et al.* (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327, 1122–1126
25. Morales, J. *et al.* (2016) Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei*. *BMC Evol. Biol.* 16, 247
26. Combe, M. *et al.* (2015) Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. *Cell Host Microbe* 18, 424–432
27. Sanjuán, R. (2017) Collective infectious units in viruses. *Trends Microbiol.* Published online March 3, 2017. http://dx.doi.org/10.1016/j.tim.2017.02.003
28. Ostrowski, E.A. *et al.* (2015) Genomic signatures of cooperation and conflict in the social amoeba. *Curr. Biol.* 25, 1661–1665
29. Robinson, G.E. *et al.* (2005) Sociogenomics: social life in molecular terms. *Nat. Rev. Genet.* 6, 257–270
30. Foster, K.R. *et al.* (2007) What can microbial genetics teach sociobiology? *Trends Genet.* 23, 74–80
31. Ghoul, M. *et al.* (2014) Toward an evolutionary definition of cheating. *Evolution* 68, 318–331
32. Jones, E.I. *et al.* (2015) Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecol. Lett.* 18, 1270–1284
33. Strassmann, J.E. and Queller, D.C. (2011) Evolution of cooperation and control of cheating in a social microbe. *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl. 2), 10855–10862
34. Strassmann, J.E. *et al.* (2000) Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* 408, 965–967
35. Ross-Gillespie, A. (2007) Frequency dependence and cooperation: theory and a test with bacteria. *Am. Nat.* 170, 331–342
36. Griffin, A.S. *et al.* (2004) Cooperation and competition in pathogenic bacteria. *Nature* 430, 1024–1027
37. Cordero, O.X. *et al.* (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20059–20064
38. Andersen, S.B. *et al.* (2015) Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10756–10761
39. Zhou, X. *et al.* (2015) Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. *Genome Biol. Evol.* 7, 2407–2416
40. Hanschen, E.R. *et al.* (2016) The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat. Commun.* 7, 11370
41. Simola, D.F. *et al.* (2013) Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235–1247
42. Laine, V.N. *et al.* (2016) Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* 7, 10474
43. Moran, N.A. *et al.* (2008) Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190
44. McAdams, H.H. and Arkin, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* 15, 65–69
45. Perez-Rueda, E. *et al.* (2004) Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. *Comput. Biol. Chem.* 28, 341–350
46. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* 136, 615–628
47. Henderson, I.R. *et al.* (1999) Molecular switches – the ON and OFF of bacterial phase variation. *Mol. Microbiol.* 33, 919–932
48. Wang, J. *et al.* (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493, 664–668
49. Purcell, J. *et al.* (2014) Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* 24, 2728–2732
50. Kupper, C. *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* 48, 79–83
51. Haskett, T.L. *et al.* (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12268–12273
52. Sachs, J.L. *et al.* (2011) Evolutionary transitions in bacterial symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl. 2), 10800–10807
53. Mc Ginty, S.E. *et al.* (2011) Horizontal gene transfer and the evolution of bacterial cooperation. *Evolution* 65, 21–32
54. Smith, J. (2001) The social evolution of bacterial pathogenesis. *Proc. Biol. Sci.* 268, 61–69
55. Nogueira, T. *et al.* (2009) Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.* 19, 1683–1691
56. Borgeaud, S. *et al.* (2015) The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science* 347, 63–67
57. Diard, M. *et al.* (2013) Stabilization of cooperative virulence by the expression of an avirulent phenotype. *Nature* 494, 353–356
58. Herrero, A. *et al.* (2016) The multicellular nature of filamentous heterocyst-forming cyanobacteria. *FEMS Microbiol. Rev.* 40, 831–854
59. Patalano, S. *et al.* (2015) Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13970–13975
60. Linksvayer, T.A. and Wade, M.J. (2016) Theoretical predictions for sociogenomic data: the effects of kin selection and sex-limited expression on the evolution of social insect genomes. *Front. Ecol. Evol.* 4, 1–10
61. Linksvayer, T.A. and Wade, M.J. (2009) Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* 63, 1685–1696
62. Warner, M.R. *et al.* (2016) Genomic signature of kin selection in an ant with obligately sterile workers. *Mol. Biol. Evol.* (in press)
63. Vojvodic, S. *et al.* (2015) The transcriptomic and evolutionary signature of social interactions regulating honey bee caste development. *Ecol. Evol.* 5, 4795–4807
64. Harpur, B.A. *et al.* (2014) Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2614–2619
65. Nielsen, H.B. *et al.* (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* 32, 822–828
66. Smith, E.E. *et al.* (2005) Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J. Bacteriol.* 187, 2138–2147
67. McInerney, J.O. *et al.* (2017) Why prokaryotes have pangenomes. *Nat. Microbiol.* 2, 17040
68. Libbrecht, R. *et al.* (2016) Robust DNA methylation in the clonal raider ant brain. *Curr. Biol.* 26, 391–395
69. Fisher, R.M. *et al.* (2013) Group formation, relatedness, and the evolution of multicellularity. *Curr. Biol.* 23, 1120–1125
70. Datta, M.S. *et al.* (2016) Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* 7, 11965
71. Myrold, D.D. and Nannipieri, P. (2014) Classical techniques versus omics approaches. In *Omics in Soil Science* (Nannipieri, P. *et al.*, eds), Caister Academic Press, pp. 179–187
72. Gardner, A. (2012) Evolution of maternal care in diploid and haplodiploid populations. *J. Evol. Biol.* 25, 1479–1486