

12-20-2019


## A Drive to Driven Model of Mapping Intraspecific Interaction Networks.

Libo Jiang  
*Beijing Forestry University*

Jian Xu  
*Chinese Academy of Fishery Sciences*

Mengmeng Sang  
*Beijing Forestry University*

Yan Zhang  
*Chinese Academy of Fishery Sciences* <https://jdc.jefferson.edu/endocrinologyfp>

 Part of the [Translational Medical Research Commons](#)  
Meixia Ye  
*Beijing Forestry University*

[Let us know how access to this document benefits you](#)

### Recommended Citation

Jiang, Libo; Xu, Jian; Sang, Mengmeng; Zhang, Yan; Ye, Meixia; Zhang, Hanyuan; Wu, Biyin; Zhu, Youxiu; Xu, Peng; Tai, Ruyu; Zhao, Zixia; Jiang, Yanliang; Dong, Chuanju; Sun, Lidan; Griffin, Christopher H.; Gragnoli, Claudia; and Wu, Rongling, "A Drive to Driven Model of Mapping Intraspecific Interaction Networks." (2019). *Division of Endocrinology, Diabetes and Metabolic Diseases Faculty Papers*. Paper 2.  
<https://jdc.jefferson.edu/endocrinologyfp/2>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Division of Endocrinology, Diabetes and Metabolic Diseases Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

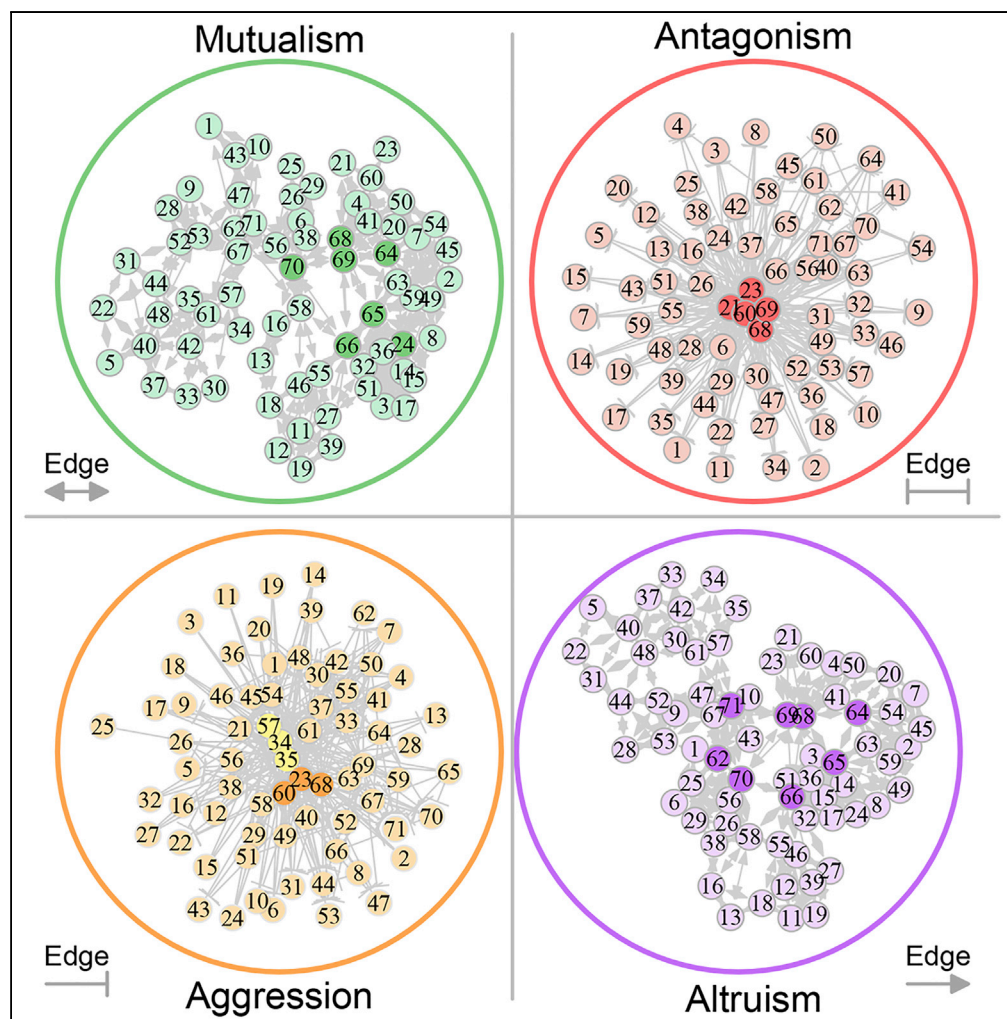
---

## Authors

Libo Jiang, Jian Xu, Mengmeng Sang, Yan Zhang, Meixia Ye, Hanyuan Zhang, Biyin Wu, Youxiu Zhu, Peng Xu, Ruyu Tai, Zixia Zhao, Yanliang Jiang, Chuanju Dong, Lidan Sun, Christopher H. Griffin, Claudia Gragnoli, and Rongling Wu

## Article

## A Drive to Driven Model of Mapping Intraspecific Interaction Networks



Libo Jiang, Jian Xu, Mengmeng Sang, ..., Christopher H. Griffin, Claudia Gragnoli, Rongling Wu

rwu@bjfu.edu.cn,  
rwu@phs.psu.edu

## HIGHLIGHTS

We develop a new theory for complex-trait mapping by integrating behavioral ecology

This theory can characterize how QTL drive cooperation or competition in populations

It can also illustrate how the activation of QTL is driven by ecological interactions

The new theory leverages interdisciplinary studies of genetics, ecology, and evolution

Jiang et al., iScience 22, 109–122  
December 20, 2019 © 2019  
The Author(s).  
<https://doi.org/10.1016/j.isci.2019.11.002>

## Article

## A Drive to Driven Model of Mapping Intraspecific Interaction Networks

Libo Jiang,<sup>1,9</sup> Jian Xu,<sup>2,9</sup> Mengmeng Sang,<sup>1</sup> Yan Zhang,<sup>2</sup> Meixia Ye,<sup>1</sup> Hanyuan Zhang,<sup>2</sup> Biyin Wu,<sup>2</sup> Youxiu Zhu,<sup>2</sup> Peng Xu,<sup>2,3</sup> Ruyu Tai,<sup>2</sup> Zixia Zhao,<sup>2</sup> Yanliang Jiang,<sup>2</sup> Chuanju Dong,<sup>2,4</sup> Lidan Sun,<sup>1</sup> Christopher H. Griffin,<sup>5</sup> Claudia Gragnoli,<sup>6,7</sup> and Rongling Wu<sup>1,8,10,\*</sup>

## SUMMARY

Community ecology theory suggests that an individual's phenotype is determined by the phenotypes of its coexisting members to the extent at which this process can shape community evolution. Here, we develop a mapping theory to identify interaction quantitative trait loci (QTL) governing inter-individual dependence. We mathematically formulate the decision-making strategy of interacting individuals. We integrate these mathematical descriptors into a statistical procedure, enabling the joint characterization of how QTL drive the strengths of ecological interactions and how the genetic architecture of QTL is driven by ecological networks. In three fish full-sib mapping experiments, we identify a set of genome-wide QTL that control a range of societal behaviors, including mutualism, altruism, aggression, and antagonism, and find that these intraspecific interactions increase the genetic variation of body mass by about 50%. We showcase how the interaction QTL can be used as editors to reconstruct and engineer new social networks for ecological communities.

## INTRODUCTION

Quantitative genetic theory has long focused on modeling how the phenotype of an individual is determined by its genes, known as quantitative trait loci (QTL), and the environment where it grows (Ritchie et al., 2015). An increasing body of evidence has revealed that an individual's phenotype in a population is also affected by the phenotypes of other members that coexist with it (Magnuson, 1962; Wolf et al., 1998; Shuster et al., 2006; Ribas et al., 2017; Schneider et al., 2017; Santostefano et al., 2017). As such, how a particular individual performs is influenced not only directly by its own QTL, but also indirectly by the QTL of its conspecifics (Jiang et al., 2018). For instance, in an association study of laying hens, a set of genes from a single hen were identified within the serotonin pathway to affect the feather condition of its cage mates (Biscarini et al., 2010). The flowering gene FRIGIDA from focal plants in *Arabidopsis* affects the developmental processes of their neighbors, according to genetic mapping using structural equation models (Wolf et al., 2011). In *Drosophila melanogaster*, several QTL detected for aggressive behavior are at play by interacting with social environments (Rohde et al., 2017).

Although inter-individual interdependence and interactions inducing phenotypic variation involve a genetic component, existing genetic mapping theory does not enable the detailed characterization of how the underlying QTL act in a mapping population. The genetic effects of QTL may be activated by ecological interactions, such as competition, where one individual grows at the cost of others exploiting the same resources, or cooperation, by which multiple individuals can better buffer against environmental perturbations than any single one alone (Fisher and Mcadam, 2017). These ecologically meaningful QTL can be better identified if we equip a mapping approach with the ecological and social principles that can explain why an individual chooses to compete or cooperate with others. The motivation of this study is to upgrade quantitative genetic theory by embedding fundamental principles of competition and cooperation, to a level at which geneticists can map specific QTL responsible for ecological interactions, estimate how these QTL affect population phenotypes through direct and indirect effects, and test how ecological interactions can induce new genetic variation for complex traits.

To test this theory, we designed and conducted a QTL mapping experiment by genotyping a full-sib family (H1) of the common carp (*Cyprinus carpio*) and culturing its  $n = 71$  siblings in a shared water pool. Previous cultural experiments showed that fish growth, behavior, and survival are highly plastic to the crowdedness of the environment (Magnuson, 1962; Fox and Flowers, 1990; Szkudlarek and Zakes, 2007; Ribas et al.,

<sup>1</sup>Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

<sup>2</sup>Key Laboratory of Aquatic Genomics, Ministry of Agriculture, CAFS Key Laboratory of Aquatic Genomics and Beijing Key Laboratory of Fishery Biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141, China

<sup>3</sup>Fujian Collaborative Innovation Center for Exploitation and Utilization of Marine Biological Resources, Xiamen University, Xiamen, Fujian 361102, China

<sup>4</sup>College of Fishery, Henan Normal University, Xinxiang, Henan 453007, China

<sup>5</sup>Applied Research Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

<sup>6</sup>Division of Endocrinology, Diabetes, and Metabolic Disease, Translational Medicine, Department of Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA 19106, USA

<sup>7</sup>Molecular Biology Laboratory, Bios Biotech Multi Diagnostic Health Center, Rome 00197, Italy

<sup>8</sup>Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033, USA

<sup>9</sup>These authors contributed equally

<sup>10</sup>Lead Contact

\*Correspondence: [rwu@bjfu.edu.cn](mailto:rwu@bjfu.edu.cn), [rwu@phs.psu.edu](mailto:rwu@phs.psu.edu)

<https://doi.org/10.1016/j.isci.2019.11.002>



2017). In a natural ecosystem of coral reef, fish make their decisions to feed on algae or escape from predators according to actions of other fish (Gil and Hein, 2017). As such, we anticipate that pervasive social interactions occur among the co-cultured fish, which exert an impact on fish phenotype. Traditional mapping approaches simply associate phenotype with genotype, without considering social interactions. The application of these approaches to our mapping population detected no QTL responsible for fish body mass (Figure S1), a trait that is sensitive to competition (Magnuson, 1962). However, when the same data were analyzed under our theory, a number of QTL have been identified. To validate these discoveries, we conducted two additional mapping experiments, from each of which consistent results are obtained.

## RESULTS

### Mathematical Descriptors of Ecological Interactions

In a socialized environment, a fish may maneuver its living territory by continuously changing its neighbors to which it pays attention (Jiang et al., 2017) so as to maximize its chance for survival and reproductive success (McFarland, 1977; Dugatkin and Reeve, 2000). This process, often guided by rational choice-based game theory (Harp, 2017), as recognized in humans (Park et al., 2017), rodents (Dias-Ferreira et al., 2009; Friedman et al., 2017), and microbes (Damore and Gore, 2012), incurs a so-called collective motion phenomenon, ubiquitous across the animal kingdom (Vicsek and Zafeiris, 2012; Jiang et al., 2017). Under natural selection, animal collective behavior has been shaped toward two tendencies. First, animals tend to swarm, flock, or shoal with individuals that resemble themselves in a cooperative way by which the so-called oddity effect, i.e., those individuals displaying difference in appearance from the group are at a greater risk to be predated (Hoare et al., 2000), can be avoided. Thus, animals of roughly similar size, color, and even smell in a population enjoy mutual cooperation and coordination (Camazine et al., 2001; Sumpter, 2006, 2010; Herbert-Read et al., 2011), and the similarity of two animals is proportional to the degree of the desire by which they cooperate. In mathematics, the similarity of two variables is positively correlated with their product, given that their sum is fixed. Taken together, we hypothesize that the product of two animals' body sizes can serve as a proxy for the strength of *mutualism*. In contrast, we use the inverse of the product of body sizes of two animals to approximately measure the strength of their antagonism.

Second, animals of larger body size tend to display agonistic behavior to those of smaller body size when a limited amount of resource needs to be allocated among members of the same population (Chance and Larson, 1976; Desjardins et al., 2012; Romenskyy et al., 2017). As an aggressive and defensive action, this behavior is adaptive, widely believed to play an important role in resource acquisition, reproductive success, and survival (Pan et al., 2010). Hence, we hypothesize that the ratio of body size of a larger over a smaller animal in the socialized environment reflects the extent to which the former exerts its *aggression* toward the latter. Accordingly, the body size difference of larger and smaller animals, divided by the body mass of the larger one, can be used as a surrogate for the strength of *altruism*. Based on the above-mentioned analysis, we derive mathematical descriptors to measure four types of intraspecific interactions, mutualism, antagonism, aggression, and altruism, by examining and comparing the body sizes of two interactive animals (Figure 1).

### Biological Justification of Interaction Measures

For a particular pair of animals in co-culture, we name the larger individual as L and the smaller individual as S. Let  $w_L$  and  $w_S$  denote the body size of L and S in co-culture, respectively. We argue that Figure 1's mathematical descriptors derived from  $w_L$  and  $w_S$  can measure the strengths of different interaction types that occur between the animals. To test these hypotheses, we analyze two real datasets, one from a cultural experiment of fish and the second from a published bacterial cultural study (Jiang et al., 2018). In each experiment, organisms were paired and two members in each pair were cultured both separately and jointly. Substantial evidence suggests that the organism often changes its phenotype in response to ecological interactions when it is shifted from an isolated environment to a socialized environment (Bohn and Amundsen, 2004; Fordyce, 2006; Lang and Benbow, 2013; Gamfeldt et al., 2013; Barraclough, 2015; Gracia-Lázaro et al., 2018). By quantifying the extent to which the phenotypic traits of the two individuals change from monoculture to co-culture, the strength of their ecological interaction can be measured and assessed.

We use  $u_L$  and  $u_S$  to denote the body size of individuals L and S in monoculture, respectively. Note that  $u_L$  is not necessarily greater than  $u_S$ , although  $w_L$  is always greater than  $w_S$  by definition. If two individuals cooperate with each other, then the relative body size of each individual in co-culture over monoculture should

		Small animal S	
		Cooperation	Competition
Large animal L	Cooperation	$z_{mu} = w_L w_S / (w_L - w_S)$	$z_{al} = (w_L - w_S) / w_L$
	Competition	$z_{ag} = w_L / w_S$	$z_{an} = 1 / [(w_L w_S)(w_L - w_S)]$

**Figure 1. Mathematical Descriptors of Four Types of Ecological Interactions, Mutualism ( $z_{mu}$ ), Antagonism ( $z_{an}$ ), Aggression ( $z_{ag}$ ), and Altruism ( $z_{al}$ )**

We use  $w_L$  and  $w_S$  to denote phenotypic values of a larger animal L and a smaller animal S, respectively, constituting a pair in a mapping population. The product of phenotypic values between two animals is used as a descriptor for the strength of mutualism, i.e., how much the two animals benefit from one another through cooperation (Zhu et al., 2016). The strength of antagonism is described by the inverse of the product of phenotypic values, reflecting how much one animal grew reciprocally at a cost of the other. To adjust the scale effect, these two descriptors are normalized by dividing them by the phenotypic difference of the larger from the smaller animal. The ratio of phenotypic values of the larger over the smaller animal is used to measure the strength of aggression, by which the former grows by harming the latter. The strength of altruism is calculated as one minus the ratio of phenotypic values of the smaller over the larger animal.

not be less than 1.0 (Ghoul and Mitri, 2016). If one individual is aggressive on the other, i.e., the former grows at a cost of the latter, then the relative body size of the former over the latter would increase when the two individuals are relocated from their respective isolated environments to the common environment. Accordingly, if one individual is altruistic toward the other, i.e., the former sacrifices itself to benefit the latter, then the relative body size of the latter in co-culture over monoculture should be larger than the relative body size of the former in co-culture over monoculture. Based on these lines of reasoning, we use  $M_u = (w_L/u_L + w_S/u_S)/2$  to quantify the strength of mutualism between individuals L and S,  $A_g = (w_L/w_S)/(u_L/u_S)$  to quantify the strength of individual L's aggression toward individual S, and  $A_l = (w_S/u_S)/(w_L/u_L)$  to quantify the strength of individual L's altruism toward individual S.

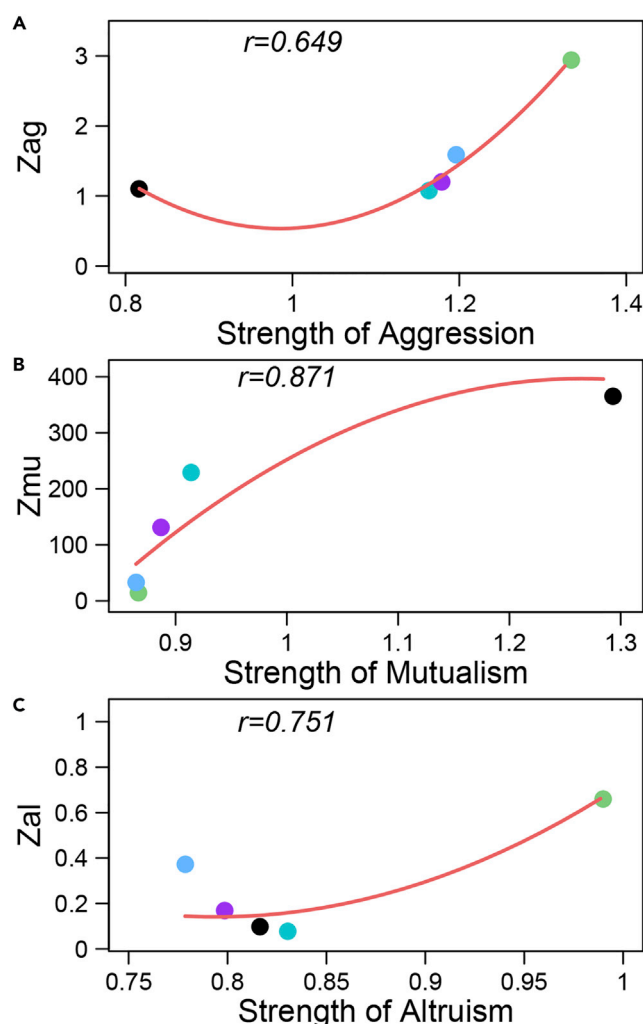
### Fish Experiment

We sampled five fish pairs from a population, in which the relative size of a smaller over larger one is 0.10, 0.38, 0.61, 0.80, and 1.00, with the larger one having a roughly similar size among pairs. Each pair was repeated four times. We reared each pair of fish in shared and isolated water buckets and measured their body mass 2 weeks after the fish was cultured. We calculated gains of body mass for each fish during culture.

Using the expressions given in Figure 1, we calculated and plotted parameters  $z_{ag}$ ,  $z_{mu}$ , and  $z_{al}$  against  $A_g$ ,  $M_u$ , and  $A_l$  for body mass gain across different fish pairs, respectively. We can test how well these three parameters can be used to measure the strengths of mutualism, parasitism, and altruism. It is interesting to find that  $z_{ag}$  is positively correlated with  $A_g$  (Figure 2A), thus suggesting that the former can approximately represent the strength of competition, especially the strength of aggression. We found that  $z_{mu}$  is positively correlated with  $M_u$  (Figure 2B), indicating that the former can well serve as a proxy to quantify the strength of mutualism. The positive correlation between  $z_{al}$  and  $A_l$  (Figure 2C) implies that the former is a good representation of the strength of altruism. From the above-mentioned analysis of fish data, it is suggested that the mathematical descriptors proposed can be used to measure different types of ecological interactions.

### Microbial Experiment

Microbes have been widely used as a system to study ecological interactions (Damore and Gore, 2012). We further validated Figure 1's mathematical descriptors by re-analyzing a published bacterial data. Jiang et al. (2018) cultured two bacterial species, *Escherichia coli* and *Staphylococcus aureus*, in socialized and socially isolated conditions, respectively. They collected 45 diverse bacterial strains from each species. Each strain from one species was grown in monoculture and its interspecific pair with a randomly selected



**Figure 2. Biological Validation of Interaction Measures in a Fish Experiment**

Scatterplots of mathematical descriptors given in Figure 1 against the strength of ecological interactions across five different pairs of fish (dots) with relative body mass 0.10, 0.38, 0.61, 0.80, and 1.00.

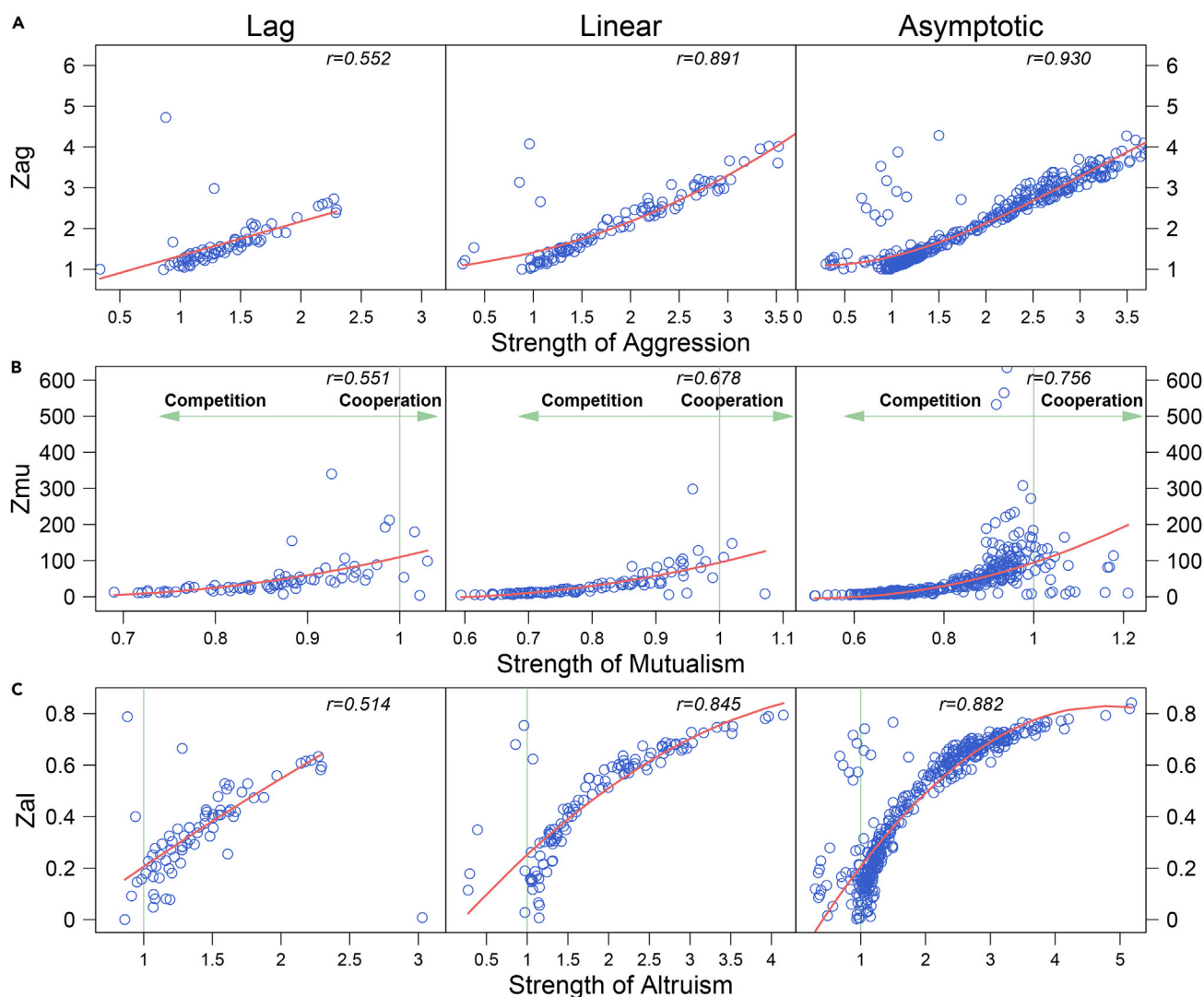
(A) Aggression descriptor ( $z_{ag}$ ) versus the strength of aggression.

(B) Mutualism descriptor ( $z_{mu}$ ) versus the strength of mutualism.

(C) Altruism descriptor ( $z_{al}$ ) versus the strength of altruism. The relationship between two variables is roughly fitted by a curve, with correlation coefficient ( $r$ ) given within each plot.

strain from the other species grown in co-culture. The abundance of each strain was measured once every 2 h during the first 24 h, followed by once every 4 h till 36 h, after the two types of culture were initiated.

Organismic growth obeys a certain rule that can be described by a growth equation (West et al., 2001). We used an optimal growth equation to fit time-dependent abundance data of each strain and further partitioned its growth curve into lag, linear, and asymptotic phases (Zwietering et al., 1990). Using the mathematical expressions of Figure 1, we calculated parameters  $z_{ag}$ ,  $z_{mu}$ , and  $z_{al}$  at each time points and plotted these parameters against  $A_g$ ,  $M_u$ , and  $A_l$ , respectively, estimated from co-culture and mono-culture data across all strains. We found that  $z_{ag}$  is positively correlated with  $A_g$  (Figure 3A) ( $p < 0.01$ ), showing the effectiveness of the former to measure the strength of aggression. These two variables display the strongest correlation at the asymptotic phase, followed by one at the linear and lag phases. This indicates that the ratio of a larger over smaller strain can better serve as a measure of the strength of aggression when the growth of strains tends to be stable. We found that  $z_{mu}$  is positively correlated with  $M_u$  ( $p < 0.01$ ), especially at the asymptotic phase (Figure 3B;  $p < 0.001$ ), suggesting that the former can



**Figure 3. Biological Validation of Interaction Measures in a Bacterial Experiment**

Scatterplots of mathematical descriptors given in Figure 1 against the strength of ecological interactions across 45 interspecific pairs of strains from *E. coli* strains and *S. aureus* at three distinct phases of microbial growth (lag, linear, and asymptotic).

(A) Aggression descriptor ( $z_{ag}$ ) versus the strength of aggression.

(B) Mutualism descriptor ( $z_{mu}$ ) versus the strength of mutualism. The strength of mutualism is measured by the average of the ratio of abundance of each bacterial species in co-culture to monoculture. Thus, this ratio average quantifies the strength of cooperation if it is above 1 and the strength of competition if it is below 1.

(C) Altruism descriptor ( $z_{al}$ ) versus the strength of altruism. Dots represent observations of different interspecific strain pairs at each time point. The relationship between two variables is roughly fitted by a curve, with correlation coefficient ( $r$ ) given within each plot.

be effectively used as the strength of cooperation. The  $z_{mu}$  values are much smaller in the competition zone ( $M_u < 1$ ) than cooperation zone ( $M_u > 1$ ). We found that  $z_{al}$  is positively correlated with  $A_i$  across strain pairs at three distinct phases, especially at linear and asymptotic phases (Figure 3C;  $p < 0.001$ ), suggesting that the former can be effectively used as a proxy to measure the strength of altruism toward a larger individual from a smaller individual.

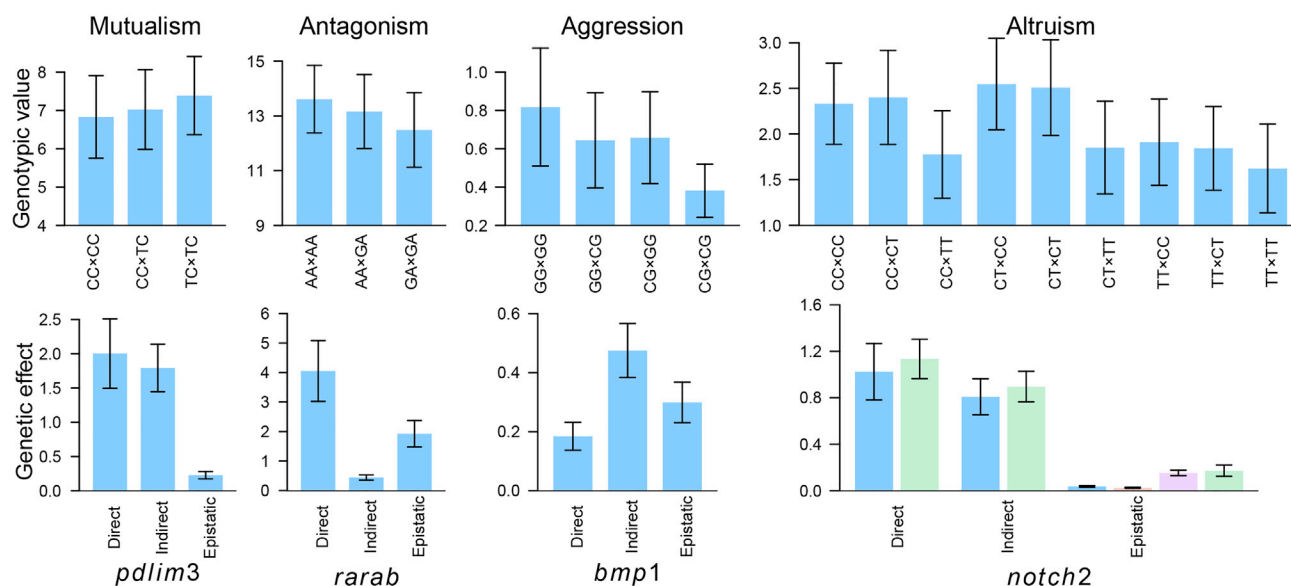
In summary, we formulate the mathematical descriptors of four typical ecological interactions, mutualism, antagonism, aggression, and altruism. We have carried out the fish experiment to validate the biological relevance of these descriptors, which was confirmed by a microbial experiment. A statistical model is implemented to map the genetic architecture of ecological interactions by treating these descriptors as phenotypes.

### Identification of Social QTL and Their Biological Relevance

The biological validation of the mathematical descriptors allows us to calculate and use four derived parameters,  $Z_{mu}$ ,  $Z_{an}$ ,  $Z_{ag}$ , and  $Z_{al}$  (Figure 1), as measures of the strength of mutualism, antagonism, aggression, and altruism, respectively, between each pair of fish in our mapping population. We used these parameters to construct the networks of each interaction type. These ecological interaction networks were incorporated into the statistical framework of QTL mapping (see the Transparent Methods). Among 39,960 high-density SNPs (with an average marker distance of 0.75 cM), our model identified 158 QTL distributed over various chromosomes for body mass, including 80 acting through mutualism, 45 through antagonism, 98 through aggression, and 76 through altruism. Yet, no QTL for body mass can be detected by traditional approaches (Figure S1). We replicated the mapping experiment by generating two independent full-sib families G1 ( $n = 115$ , with 97,532 SNPs) and Z22 ( $n = 62$ , with 86,370 SNPs) from different common carp parents, from each of which a similar conclusion was reached; i.e., a number of QTL on different regions of the genome were detected by the ecological interaction-implemented model, whereas none was detected by traditional approaches (Figures S2 and S3).

We performed an extensive gene enrichment analysis for the significant SNPs by screening their up- and down-stream 10 kb regions on the sequenced genome of the common carp (Xu et al., 2014). Together, a large proportion of QTL detected from three mapping families were annotated to candidate genes: 86.2% for mutualism, 85.7% for antagonism, 85.9% for aggression, and 85.4% for altruism (Tables S1, S2, S3, and S4). We found that significant SNPs located in clusters are individually annotated to different genes (Figures S1–S3). All candidate genes have been previously reported in association with growth traits. For example, *pdlim4* (PDZ and LIM domain protein 4) of QTL carp227526 from family Z22 and *pdlim3* of QTL carp168806 from family G1 detected simultaneously by mutualistic, aggressive, and altruistic models are closely related to muscle growth and development (Hsieh et al., 2014). These three models also detected *gpc4* gene of carp028224 from family G1, *notch2* gene of QTL carp152585 from family Z22, and *gpr101* gene of QTL carp123609 from family G1. A family of *gpc* genes, e.g., *gpc1a*, *gpc3*, and *gpc4*, encoding glypicans, are expressed during the gastrulation stage of zebrafish, with their expression becoming more tissue specific and defined at the somitogenesis stages (Gupta and Brand, 2013). *notch2* has been widely reported to play a vital role in skeletal and muscle development (Zanotti and Canalis, 2013). *gpr101* gene of QTL carp123609 from family G1 is involved in skeletal development (Beckers et al., 2015), and its other close GPR family members interact with IGFs and are crucial for muscle and body growth (Yang et al., 2014). Additionally, other genes identified uniquely by a certain model are also relevant in terms of biological functions; for instance, the genes *prss23* of carp170891 from family G1, *rarab* of carp055558 from family H1, *bmp1* of carp017510 from family H1, and *acer1* of carp117856 from family H1 were detected by the mutualistic, antagonistic, aggressive, and altruistic models, respectively. Molecular experiments in zebrafish showed that *prss23* was essential for endothelial-to-mesenchymal transition during valvulogenesis (Chen et al., 2013). Mice studies showed the involvement of *rarab* in fatty acid oxidation and energy homeostasis (Li et al., 2013). *bmp1* (bone morphogenetic protein 1) affects embryo development and osteogenesis (Muir et al., 2014) and is essential for human type 1 collagen fibrillogenesis (Valencia et al., 2014). *acer1* is important for mammalian skin homeostasis and the regulation of energy expenditure (Liakath-Ali et al., 2016).

To glean insight into the genetic mechanisms underlying the formation of body mass, we further performed GO and KEGG enrichment analyses for the QTL detected (Tables S5, S6, S7, and S8, Figures S4–S6). GO analysis identified significant enrichments of mutualism, aggression, and altruism QTL in “multicellular organism development (GO: 0007275)” and “fin development (GO: 003333),” both of which include two genes reported to affect zebrafish development, *notch2* (Zanotti and Canalis, 2013) and *hmcn1* (Feitosa et al., 2012). GO terms were enriched by the mutualistic model in “regulation of Notch signaling pathway (GO: 0008593),” which plays a vital role in bone and neurite development (Zanotti and Canalis, 2013). The antagonist model enriched “steroid hormone mediated signaling pathway (GO: 0043401)” (Li et al., 2013) and “B cell activation (GO: 0042113).” The enriched “B cell activation” suggests that stress-related genes, such as *prkcbb*, participate in fish-fish competition by regulating the  $D_2$ -like dopamine autoreceptor (Luderman et al., 2015). “Somitogenesis (GO: 0001756),” uniquely identified by the aggressive model, is interestingly closely related to myogenesis and muscle growth (Gupta and Brand, 2013), which enhance the fish to develop a strong capacity for aggression. “Lipid metabolic process (GO: 0006629),” only detected by the altruistic model, is remarkably involved in energy expenditure (Liakath-Ali et al., 2016) and inhibits aggression, invoking altruism.



**Figure 4. Quantitative Genetic Dissection of Genotype Combination Values For Fish Body Mass**

**Upper panel:** Genotypic values of combinations CC × CC, CC × TC, and TC × TC at *pdlim3* (testcross QTL) for the strength of mutualism; combinations AA × AA, AA × GA, and GA × GA at *thraa* (testcross QTL) for the strength of antagonism; combinations GG × GG, GG × CG, CG × GG, and CG × CG at *bmp1* (testcross QTL) for the strength of aggression; and combinations CC × CC, CC × CT, CC × TT, CT × CC, CT × CT, CT × TT, TT × CC, TT × CT, and TT × TT at *notch2* (intercross QTL) for the strength of altruism. **Lower panel:** Direct genetic effects that describe how the alleles of a fish in a pair affects its own body mass; indirect genetic effects that specify how each fish gene affects its conspecific's phenotypes; and genome-genome epistatic effects that quantify how the interactions between genes of two fish affect the phenotype of each fish. For the intercross QTL, both the direct and indirect effects include additive (blue) and dominant (green) effects and genome-genome epistatic effects include additive × additive, additive × dominant, dominant × additive, and dominant × dominant effects (in order from left to right). Standard errors for each value are given.

KEGG analysis found even more fascinating enriched pathways (Table S9). The mutualistic, aggressive, and altruistic models enriched four pathways closely associated with body weight: the “neuroactive ligand-receptor interaction,” “mTOR signaling pathway,” “progesterone-mediated oocyte maturation,” and “adrenergic signaling in cardiomyocytes.” For example, *gnai3* in the last pathway has been reported to regulate pig postnatal growth by engaging in miRNA-mRNA interactions (Ye et al., 2015). Mutualistic and altruistic models both identified the “Wnt signaling pathway,” which plays an important role in body axis patterning, cell proliferation, and cell migration and, therefore, embryonic development. These processes within the Wnt signaling pathway not only are necessary for bone and muscle formation but also control adult bone marrow, skin, and intestine tissue regeneration (Clevers et al., 2014), which is key to longevity and function.

### How QTL Act: Direct, Indirect, and Genome-Genome Epistatic Effects

Our theory can partition the genotypic value of a social QTL into its different genetic components (see the Methods). *pdlim3* detected from family G1 is a testcross QTL for mutualism with two genotypes paired among the fish. The fish carrying the same genotype TC at this mutualism QTL are more cooperative with each other than with those carrying the alternative CC (Figure 4). At the *rarab* gene detected from family H1, stronger antagonism occurs between the fish of the same genotype AA than between those carrying different genotypes, and the fish with the same alternative genotype GA are the least antagonistic to each other. The fish carrying GG at *bmp1*, detected from family H1, repress those with the same genotype much more severely than with the alternative CG, whereas the fish of the same genotype CG are the least aggressive to each other. As an intercross QTL, the *notch2* detected from family H1, G1, and Z22, have three genotypes (CC, CT, and TT) forming nine genotype combinations among pairing fish. Genotype CT is more altruistic to the same genotype and genotype CC than to genotype TT, and genotype TT is the least altruistic to the same genotype among all combinations. Our model can separate the direct genetic effect of a QTL from one fish on its own body mass; the indirect genetic effect of a QTL from one fish on the body mass of its pairing partner; and the genome-genome epistatic effect of a QTL from two fish on the body mass of each fish. We found that mutualism *pdlim3* controls the body mass of fish not only through its direct effect

but also through its indirect effect (Figure 4). The influence of genome-genome epistatic effect was evidently detected for antagonism *rarab*. Surprisingly, indirect and genome-genome epistatic effects are more pronounced than direct effect at aggression *bmp1*. As an intercross QTL, altruism *notch2* may exert its genetic impact by additive and dominant effects and their epistatic interactions. In fact, a remarkable indirect effect through both additive and dominant inheritance triggered by this QTL was found, although its genome-genome epistatic effects are not significant.

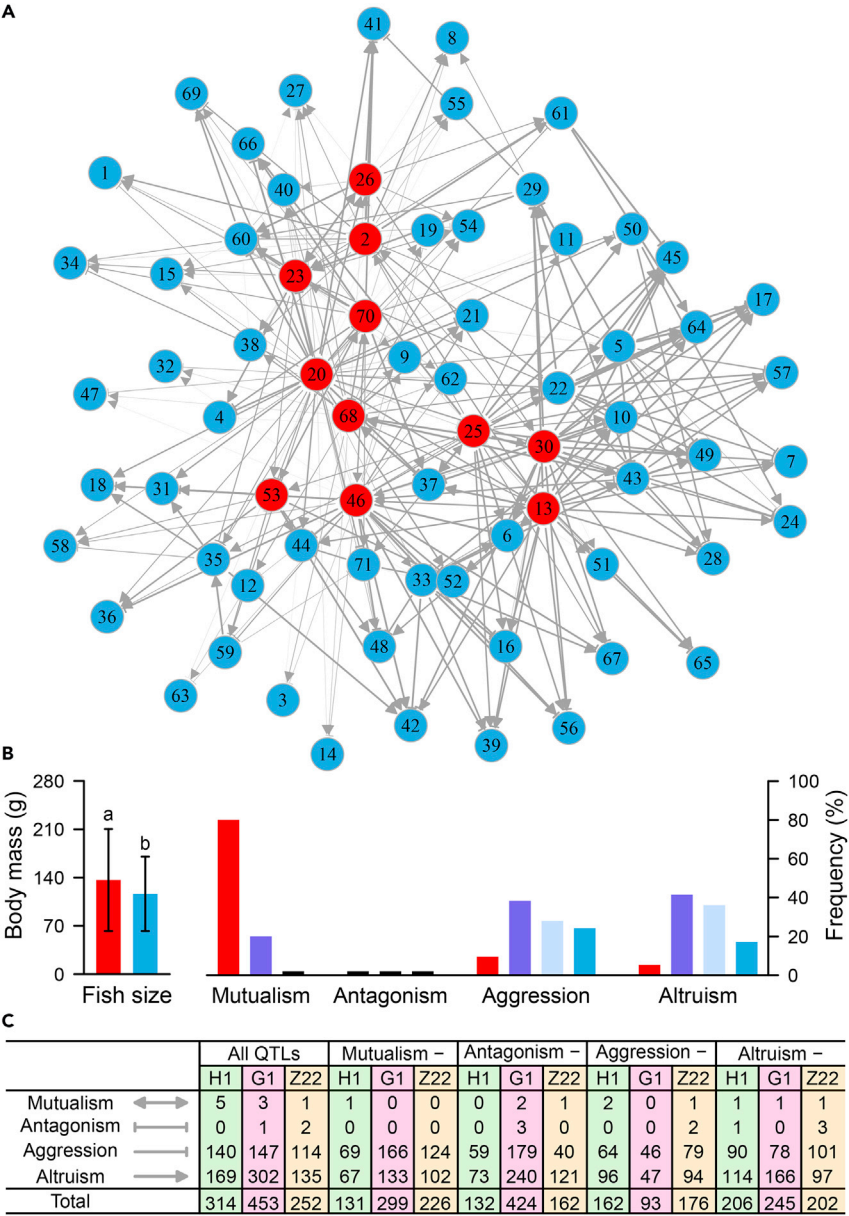
We further estimated the proportions of variance due to each of these effects to the total genetic variance at each QTL. Averaged over all QTL, indirect and genome-genome epistatic effects together explained approximately 50% of the total genetic variance for body mass, a phenomenon detected consistently in three mapping families (Table S10). These two portions of genetic components, largely neglected in previous quantitative and evolutionary genetic studies, may help geneticists chart a more complete genetic signature.

### Social Networks and QTL Networks

Using marginal genotypic values at each QTL, we modified an ordinary differential equation method (see the Methods) to infer a directed, signed, and weighted network of social interactions among the fish based on all QTL for mutualism, antagonism, aggression, and altruism. In the family H1 of 71 fish, this QTL-driven social network is composed of a total of 314 pillar connections from 2,485 possible links (Figure 5A), by which one fish connects and interacts with other fish selectively according to the game theory. For example, the network is dominated by 11 hub fish, which are larger than their marginal counterparts ( $p < 0.01$ ) (Figure 5B). Of all mutualistic relationships, 80% occurs between the hub fish, 20% between the hub and marginal fish, and none between the marginal fish. The hub fish are less aggressive toward each other than toward the marginal conspecifics, although the marginal fish have some degree of aggression toward the hub fish and other marginal fish. The hub fish are also much less altruistic toward each other, compared with how much benefit they offer to the marginal ones. Similarly, the marginal fish are less altruistic toward each other than toward the hubs, although this difference is much more moderate compared with the difference detected in the hubs. All of these fish behaviors, which are consistent with the predictions from the game theory, suggest that animal's decision making in a socialized environment involves a strong genetic component.

In this study, we investigate how the underlying QTL govern behaviors of fish-fish interactions. We reconstructed four QTL-driven social fish networks by excluding either QTL for mutualism, or antagonism, or aggression, or altruism. The number of connections within each of these networks was, respectively, reduced sharply to 137, 132, 162, and 206 (Figure 5C), suggesting that a large number of QTL are essential for the maintenance of complex social networks. Specifically, when mutualism QTL were excluded, the number of mutualistic relationships was reduced to one, compared with five in the network constructed from all detected QTL. Similarly, aggressive relationships within the aggression QTL-excluded network and altruistic relationships within the altruism QTL-excluded network both become much less frequent (i.e., 64 and 114, respectively, compared with 140 and 169 within the network from all QTL). Similar findings have been confirmed in the other two families G1 and Z22 (Figure 5C). These results suggest that mutualism, aggression, and altruism QTL play an important role in forming and preserving, respectively, mutualistic, aggressive, and altruistic relationships in an interactive community. In other words, community structure, organization, and even function can be altered, modified, and engineered by activating, repressing, or removing the expression of specific social QTL.

To demonstrate how the detected QTL jointly affect the fish social network, we implemented ordinal Bayesian networks (see the Transparent Methods) to construct a directed acyclic graph (DAG) of QTL interactions for family H1 (Figure 6). We found that mutualism and antagonism QTL that determine two extreme patterns of social behaviors organize into distinct modules, connected via aggression and altruism QTL. A total of 10 QTL (*COX5B*, *STAR*, *ADAM9*, *LMO41*, *lqsec2*, *Colgalt2*, *GPR160*, *Tnik*, *rps6ka6*, and *Msn*) pleiotropically affected the behavior of mutualism, aggression, and altruism. Other pleiotropic QTL included *VPS13A* for mutualism and aggression; *MYO1F* for mutualism and altruism; and *BBOF1*, *ODO1*, *RIFK*, *SAL*, and *AGRD1* for aggression and altruism. No QTL were detected to be shared for antagonism and the other types of interactions. Eleven QTL established a set of hub genes that modulate the structure and organization of the QTL network by activating or inhibiting other QTL. *bmp1* is socially an aggression QTL, but it is not genetically "aggressive" because its expression needs to be regulated by many other

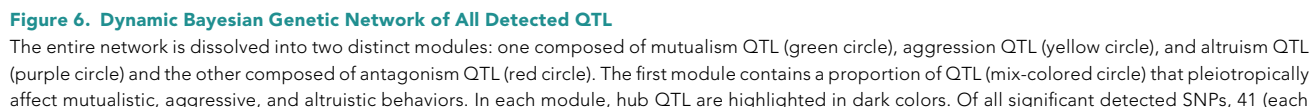


**Figure 5. A Bidirectional, Signed, and Weighted Social Network of All Fish Driven by Various Types of QTL Constructed from Ordinary Differential Equations**

(A) Social network of family H1 constructed from all QTL with edges representing how one fish interacts with others through mutualism (doubly arrowed), antagonism (doubly T-shaped), aggression (singly T-shaped), or altruism (singly arrowed). Hubs of the network are highlighted in red.

(B) The network is characterized by the difference in body mass between groups of hubs (red) and non-hubs (blue), the percentages of mutualistic and antagonistic edges among hubs (red), among hubs and non-hubs (purple), and among non-hubs, and the percentages of aggressive and altruistic edges from one fish to the second both from the hub group (red), from one fish from the hub group to the second from the non-hub group (purple), from one fish from the non-hub group to the second from the hub group (gray), and from one fish to the second both from the non-hub group (blue).

(C) The numbers of mutualistic, antagonistic, aggressive, or altruistic edges with the social networks constructed from all QTL as well as from all QTL, except for, respectively, mutualism, antagonism, aggression, and altruism QTL. Comparisons of edge numbers are given not only for family H1, but also for the two family replicates G1 and Z22.



**Figure 6. Continued**

labeled by a number) were identified as uniquely segregating in the mapped population, which was used for QTL network construction. It is possible that different uniquely segregating SNPs may correspond to the same candidate gene if they are physically close enough on the common carp genome. Candidate genes adjacent to significant SNPs are listed below. Aggression *bmp1*, mutualism *hmcn1*, and antagonism *rarab* are annotated, respectively, by SNPs #37 and #38 and SNPs #35 and #24. The arrow denotes the direction by which one gene regulates the other.

QTL. *hmcn1* affects fish mutualistic behavior, but its effect depends on the joint regulation of other QTL. Antagonism *rarab* is regulated by other genes, such as *suv420h2* and *prkcbb*, but it also modulates the expression of other genes. Overall, this QTL network helps to maintain the balance of social interactions by guiding the decision of individual fish to cooperate or compete with their conspecifics. Taken together, a detailed portrait of QTL DAG provides a mechanistic understanding of how QTL determine body mass in a fish population through their epistatic network. A similar phenomenon was also detected in families G1 and Z22, in which QTL form different but connected genetic modules according to their social behavior. These results, drawn consistently from three independent fish families, could provide evidence about the biological relevance of our theory.

**Monte Carlo Simulation**

We examined the statistical behavior of our model through computer simulation. Our model estimates reasonably well the genetic effects of QTL, including direct, indirect, and genome-genome epistatic effects, and possesses good power for QTL detection (Table S11). This can be attributed to the increase of information from pairwise phenotypes under the design of our model. The same data were analyzed by a traditional model, which shows reduced power for QTL detection. The advantage of our model is more evident when the heritability and/or sample size are modest. The false-positive rates of our model are reasonably low (<0.08) even when the mapping population is modest (e.g., 70).

**DISCUSSION**

No organism can live in absolute isolation, rather the phenotype and fitness of an organism should be determined not only by its own intrinsic properties, but also by the strategies its conspecifics develop and use in response to the biological environment (Magnuson, 1962; Ribas et al., 2017; Schneider et al., 2017; Santostefano et al., 2017). However, measuring the strength of such ecological and social interactions from a mapping experiment is highly challenging. Based on animal behavioral ecology theory, we formulate the mathematical rule of thumb to quantitatively describe the strength of different interaction types that take place in a mapping population. The cultural experiments of fish and bacteria consistently support the biological relevance of our mathematical descriptors. We propose a mapping theory for complex traits by incorporating the mathematical descriptors of ecological interactions. We further arm our theory with a computational toolkit to map and identify QTL acting through direct genetic effects (by which an individual's QTL affects its own phenotype), indirect genetic effects (by which an individual's QTL influences the phenotype of its conspecifics), and trans-genome epistatic effects (by which the interaction of QTL derived from different individuals controls each of their phenotypes).

Our theory was used in three independent mapping experiments of fish, obtaining consistent results. We estimate the contributions of direct, indirect, and trans-genome epistatic genetic effects to quantitative genetic variation and find that the latter two effects can together account for approximately half of the total genetic variance in body mass. Many earlier studies have recognized the importance of indirect genetic effects (Schneider et al., 2017; Santostefano et al., 2017), but quantification of how they contribute to genetic variation has been lacking. Our mapping theory opens a gateway to capturing these overlooked sources of genetic variation, thereby portraying a more comprehensive genetic architecture of complex traits.

Apart from its increasing precision of trait mapping, our theory raises two key interdisciplinary questions for future research. First, quantitative genetic theory has been increasingly coupled with behavioral ecology to reveal the genetic mechanisms underlying social traits, such as aggression and response to social opponents (Dingemans and Araya-Ajoy, 2015), and to uncover why selection maintains behavioral variation rather than eroding it (Santostefano et al., 2017). The major social interaction types of mutualism, antagonism, aggression, and altruism profoundly impact the structure and function of ecological communities in their unique ways. We found that these interactive processes have distinct genetic bases for fitness-related body size in the fish. To establish a complex social network, more QTL should be activated by playing a single or multifaceted role. By excluding mutualism, aggression, and altruism QTL, the fish become,

respectively, less cooperative, aggressive, and altruistic in the population. This result has an immediate implication for the genetic study and possible manipulation in the real world of behavioral variation and evolution. By repressing or even eliminating the expression of aggression QTL through modern gene editing, such as CRISPR, researchers in ecology, breeding, or medicine can create and preserve more cooperative (e.g., for the gut microbiota) or more antagonistic (e.g., for intra-tumoral cells) communities beneficial to humans.

Second, indirect genetic effects arising from communal interactions are regarded as a source of additional genetic variation, whose impact on the social-traits evolutionary dynamics, by enhancing rapid selection responses or functioning as evolutionary constraint on phenotypes, has been well documented in many experimental studies (Wolf et al., 1998; Shuster et al., 2006; Wilson et al., 2011; Schneider et al., 2017; Santostefano et al., 2017). Not only are behavioral traits affected by indirect genetic effects, but also, as shown by our result, morphological traits, such as body mass, are influenced by an indirect genetic component. Our finding is innovative and insightful; for instance, we can infer through psychology that our human behavior responds indirectly to the presence of other surrounding humans' genes and their related-effects as these both affect our psyche and choices in food, which in turn then affect our body mass. The incorporation into evolutionary studies of these indirect genetic effects and trans-genome epistatic effects, expressed at specific QTL levels, can improve our insight into how social interactions between conspecifics impose a diverse array of selective pressures on various behaviors and how evolutionary stasis occurs for phenotypic traits involved in social interactions.

### Limitations of the Study

We propose a mapping theory for charting a more complete map of the genetic architecture of complex traits by incorporating the impact of ecological interactions on phenotypic variation. Although this theory has successfully identified the previously unknown genetic variation of body size in animals, it is unclear how it works to study and dissect other types of phenotypic traits, such as disease-related and physiological traits. Furthermore, our biological justification of interaction descriptors was based on cultural experiments of mobile animals and microbes, but we do not know whether this justification can be extended to immobile plants that communicate with each other differently from the way mobile organisms do. The unique feature of our theory is to take advantage of behavioral ecology to enhance the efficiency of trait mapping. The biological processes of how different organisms cooperate or compete for living resources in populations, communities, or ecosystems are also governed by evolutionary principles, developmental biology, habitat ecology, and network science. The seamless integration of all these disciplines into our mapping theory will certainly facilitate its widespread use to construct mechanistic links from genotype to phenotype.

### METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

### DATA AND CODE AVAILABILITY

All data and computer code are given in GitHub <https://github.com/LiboJiang/Fish> or can be directly requested from the corresponding author.

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2019.11.002>.

### ACKNOWLEDGMENTS

We thank Jennifer Wilson for her editorial modification to this manuscript. This work is supported by Fundamental Research Funds for the Central Universities 2015ZCQ-SW-06 and BLX2015-23 (R.W. and M.Y.), grant 31700576 from National Natural Science Foundation of China (L.J.), grant 31600536 from National Natural Science Foundation of China (M.Y.), grant NICHD 5R01HD086911-02 from the National Institute of Health (C.G.), grant CMMI-1463482 from National Science Foundation (C.H.G.), The National Key Research and Development Program (2018YFD0900102), The National Science Foundation of China (31502151), Central Public-interest Scientific Institute Basal Research Fund, CAFS (NO. 2016GH02 and NO.2015C007), and The State Administration of Forestry of China (201404102).

## AUTHOR CONTRIBUTIONS

L.J. derived the model and conducted data analysis and computer simulation. J.X. performed the experiment, collected the data, conducted gene annotation analysis, and contributed to the writing of the Results and Material part. M.S., M.Y., and L.S. participated in model derivations, data analysis, and result interpretation. Y.Z., H.Z., B.W., Y.Z., P.X., R.T., Z.Z., Y.J., and C.D. participated in fish data collection. C.H.G. interpreted game theory and formulated its mathematical analysis. C.G. critically read and revised the manuscript. R.W. conceived the idea, supervised the overall study, and wrote the paper.

## DECLARATION OF INTERESTS

The authors have no competing interests.

Received: June 12, 2019

Revised: September 23, 2019

Accepted: November 1, 2019

Published: December 20, 2019

## REFERENCES

- Barracough, T.G. (2015). How do species interactions affect evolutionary dynamics across whole communities? *Ann. Rev. Ecol. Evol. Syst.* 46, 25–48.
- Beckers, A., Lodish, M.B., Trivellin, G., Rostomyan, L., Lee, M., Fauz, F.R., Yuan, B., Choong, C.S., Caberg, J.H., Verrua, E., et al. (2015). X-linked acrogigantism syndrome: clinical profile and therapeutic responses. *Endocr. Relat. Cancer* 22, 353–367.
- Biscarini, F., Bovenhuis, H., van Arendonk, J.A., Parmentier, H.K., Jungerius, A.P., and van der Poel, J.J. (2010). Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet.* 41, 26–38.
- Bohn, T., and Amundsen, P.-A. (2004). Ecological interactions and evolution: forgotten parts of biodiversity? *BioScience* 54, 804–805.
- Camazine, S., Deneubourg, J.-L., Franks, N., Sneyd, J., Theraulaz, G., and Bonabeau, E. (2001). *Self-Organization in Biological Systems* (Princeton University Press).
- Chance, M.R.A., and Larson, R.R. (1976). *The Social Structure of Attention* (John Wiley & Sons).
- Chen, I.H., Wang, H.H., Hsieh, Y.S., Huang, W.C., Yeh, H.I., and Chuang, Y.J. (2013). PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish. *Cardiovasc. Res.* 97, 443–453.
- Clevers, H., Loh, K.M., and Nusse, R. (2014). An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* 346, 1248012.
- Damore, J.A., and Gore, J. (2012). Understanding microbial cooperation. *J. Theor. Biol.* 299, 31–41.
- Desjardins, J.K., Hofmann, H.A., and Fernald, R.D. (2012). Social context influences aggressive and courtship behavior in a cichlid fish. *PLoS One* 7, e32781.
- Dias-Ferreira, E., Sousa, J.C., Melo, I., Morgado, P., Mesquita, A.R., Cerqueira, J.J., Costa, R.M., and Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* 325, 621–625.
- Dingemanse, D.J., and Araya-Ajoy, Y.G. (2015). Interacting personalities: behavioural ecology meets quantitative genetics. *Trends Ecol. Evol.* 30, 88–97.
- Dugatkin, L.A., and Reeve, H.K. (2000). *Game Theory and Animal Behavior* (Oxford University Press).
- Feitosa, N.M., Zhang, J., Carney, T.J., Metzger, M., Korzh, V., Bloch, W., and Hammerschmidt, M. (2012). Hemicentin 2 and Fibulin 1 are required for epidermal–dermal junction formation and fin mesenchymal cell migration during zebrafish development. *Dev. Biol.* 369, 235–248.
- Fisher, D.N., and Mcadam, A.G. (2017). Social traits, social networks and evolutionary biology. *J. Evol. Biol.* 30, 2088–2103.
- Fordyce, J.A. (2006). The evolutionary consequences of ecological interactions mediated through phenotypic plasticity. *J. Exp. Biol.* 209 (Pt 12), 2377–2383.
- Fox, M.G., and Flowers, D.D. (1990). Effect of fish density on growth, survival, and food consumption by juvenile Walleyes in rearing ponds. *Trans. Am. Fish. Soc.* 119, 112–121.
- Friedman, A., Homma, D., Bloem, B., Gibb, L.G., Amemori, K.I., Hu, D., Delcasso, S., Truong, T.F., Yang, J., Hood, A.S., and Mikofalvy, K.A. (2017). Chronic stress alters striosome-circuit dynamics, leading to aberrant decision-making. *Cell* 171, 1191–1205.
- Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D., et al. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* 4, 1340.
- Ghoul, M., and Mitri, S. (2016). The ecology and evolution of microbial competition. *Trends Microbiol.* 24, 833–845.
- Gil, M.A., and Hein, A.M. (2017). Social interactions among grazing reef fish drive material flux in a coral reef ecosystem. *Proc. Natl. Acad. Sci. U S A* 114, 4703–4708.
- Gracia-Lázaro, C., Hernández, L., Borge-Holthoefer, J., and Moreno, Y. (2018). The joint influence of competition and mutualism on the biodiversity of mutualistic ecosystems. *Sci. Rep.* 8, 9253.
- Gupta, M., and Brand, M. (2013). Identification and expression analysis of zebrafish glypicans during embryonic development. *PLoS One* 8, e80824.
- Harp, R. (2017). Collective action and rational choice explanations. *J. Philos. Res.* 42, 149–176.
- Herbert-Read, J.E., Perna, A., Mann, R.P., Schaerf, T.M., Sumpter, D.J., and Ward, A.J. (2011). Inferring the rules of interaction of shoaling fish. *Proc. Natl. Acad. Sci. U S A* 108, 18726–18731.
- Hoare, D.J., Krause, J., Peuhkuri, N., and Godin, J.G.J. (2000). Body size and shoaling in fish. *J. Fish. Biol.* 57, 1351–1366.
- Hsieh, C.L., Liu, H., Huang, Y., Kang, L., Chen, H.W., Chen, Y.T., Wee, Y.R., Chen, S.J., and Tan, B.C. (2014). ADAR1 deaminase contributes to scheduled skeletal myogenesis progression via stage-specific functions. *Cell Death Differ.* 21, 707–719.
- Jiang, L., Giuggioli, L., Perna, A., Escobedo, R., Lecheval, V., Sire, C., Han, Z., and Theraulaz, G. (2017). Identifying influential neighbors in animal flocking. *PLoS Comput. Biol.* 13, e1005822.
- Jiang, L.B., He, X.Q., Jin, Y., Ye, M.X., Sang, M.M., Chen, N., Zhu, J., Zhang, Z.R., Li, J.T., and Wu, R.L. (2018). A mapping framework of collaboration-competition QTLs that drive community dynamics. *Nat. Commun.* 9, 3010.
- Lang, J.M., and Benbow, M.E. (2013). Species interactions and competition. *Nat. Ed. Knowl.* 4, 8.
- Li, Y., Wong, K., Walsh, K., Gao, B., and Zang, M. (2013). Retinoic acid receptor  $\beta$  stimulates hepatic induction of fibroblast growth factor 21 to promote fatty acid oxidation and control whole-

body energy homeostasis in mice. *J. Biol. Chem.* 288, 10490–10504.

Liakath-Ali, K., Vancollie, V.E., Lelliott, C.J., Speak, A.O., Lafont, D., Protheroe, H.J., Ingvorsen, C., Galli, A., Green, A., Gleeson, D., and Ryder, E. (2016). Alkaline ceramidase 1 is essential for mammalian skin homeostasis and regulating whole-body energy expenditure. *J. Pathol.* 239, 374–383.

Luderman, K.D., Chen, R., Ferris, M.J., Jones, S.R., and Gnegy, M.E. (2015). Protein kinase C beta regulates the D2-Like dopamine autoreceptor. *Neuropharmacology* 89, 335–341.

Magnuson, J.J. (1962). An analysis of aggressive behavior, growth, and competition for food and space in medaka (*Oryzias latipes*) (Pisces, Cyprinodontidae). *Can. J. Zool.* 40, 313–363.

McFarland, D.J. (1977). Decision making in animals. *Nature* 269, 15–21.

Muir, A.M., Ren, Y., Butz, D.H., Davis, N.A., Blank, R.D., Birk, D.E., Lee, S.J., Rowe, D., Feng, J.Q., and Greenspan, D.S. (2014). Induced ablation of Bmp1 and Tll1 produces osteogenesis imperfecta in mice. *Hum. Mol. Genet.* 23, 3085–3101.

Pan, Y., Xu, L., Young, K.A., Wang, Z., and Zhang, Z. (2010). Agonistic encounters and brain activation in dominant and subordinate male greater long-tailed hamsters. *Horm. Behav.* 58, 478–484.

Park, H., Lee, D., and Chey, J. (2017). Stress enhances model-free reinforcement learning only after negative outcome. *PLoS One* 12, e0180588.

Ribas, L., Valdivieso, A., Díaz, N., and Piferrer, F. (2017). Appropriate rearing density in domesticated zebrafish to avoid masculinization: links with the stress response. *J. Exp. Biol.* 220, 1056–1064.

Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A., and Kim, D. (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* 16, 85–97.

Rohde, P.D., Gaertner, B., Ward, K., Sørensen, P., and Mackay, T.F.C. (2017). Genomic analysis of

genotype-by-social environment interaction for *Drosophila melanogaster* aggressive behavior. *Genetics* 206, 1969–1984.

Romensky, M., Herbert-Read, J.E., Ward, A.J.W., and Sumpter, D.J.T. (2017). Body size affects the strength of social interactions and spatial organization of a schooling fish (*Pseudomugil signifer*). *R. Soc. Open Sci.* 4, 161056.

Santostefano, F., Wilson, A.J., Niemelä, P.T., and Dingemanse, N.J. (2017). Indirect genetic effects: a key component of the genetic architecture of behaviour. *Sci. Rep.* 7, 10235.

Schneider, J., Atallah, J., and Joel, D.L. (2017). Social structure and indirect genetic effects: genetics of social behaviour. *Biol. Rev.* 92, 1027–1038.

Shuster, S.M., Lonsdorf, E.V., Wimp, G.M., Bailey, J.K., and Whitham, T.G. (2006). Community heritability measures the evolutionary consequences of indirect genetic effects on community structure. *Evolution* 60, 991–1003.

Sumpter, D.J. (2006). The principles of collective animal behaviour. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 5–22.

Sumpter, D.J. (2010). *Collective Animal Behavior* (Princeton University Press).

Szkudlarek, M., and Zakes, Z. (2007). Effect of stocking density on survival and growth performance of pikeperch, *Sander lucioperca* (L.), larvae under controlled conditions. *Aquaculture Int.* 15, 67–81.

Valencia, M., Caparrós-Martin, J.A., Sirerol-Piquer, M.S., García-Verdugo, J.M., Martínez-Glez, V., Lapunzina, P., Temtamy, S., Aglan, M., Lund, A.M., Nikkels, P.G., et al. (2014). Report of a newly identified patient with mutations in BMP1 and underlying pathogenetic aspects. *Am. J. Med. Genet. A* 164, 1143–1150.

Vicsek, T., and Zafeiris, A. (2012). Collective motion. *Phys. Rep.* 517, 71–140.

West, G.B., Brown, J.H., and Enquist, B.J. (2001). A general model for ontogenetic growth. *Nature* 413, 628–631.

Wilson, A.J., Morrissey, M.B., Adams, M.J., Walling, C.A., Guinness, F.E., Pemberton, J.M., Clutton-Brock, T.H., and Kruuk, L.E.B. (2011). Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, *Cervus Elaphus*. *J. Evol. Biol.* 24, 772–783.

Wolf, J.B., Brodie, E.D., III, Cheverud, J.M., Moore, A.J., and Wade, M.J. (1998). Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* 13, 64–69.

Wolf, J.B., Mutic, J.J., and Kover, P.X. (2011). Functional genetics of intraspecific ecological interactions in *Arabidopsis thaliana*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1358–1367.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., et al. (2014). Genome sequence and genetic diversity of the common carp. *Nat. Genet.* 46, 1212–1219.

Yang, L., Sun, C., and Li, W. (2014). Neuropeptide B in Nile tilapia *Oreochromis niloticus*: molecular cloning and its effects on the regulation of food intake and mRNA expression of growth hormone and prolactin. *Gen. Comp. Endocrinol.* 200, 27–34.

Ye, R.S., Li, M., Qi, Q.E., Cheng, X., Chen, T., Li, C.Y., Wang, S.B., Shu, G., Wang, L.N., Zhu, X.T., et al. (2015). Comparative anterior pituitary miRNA and mRNA expression profiles of Bama Minipigs and Landrace Pigs reveal potential molecular network involved in animal postnatal growth. *PLoS One* 10, e0131987.

Zanotti, S., and Canalís, E. (2013). Notch signaling in skeletal health and disease. *Eur. J. Endocrinol.* 168, R95–R103.

Zhu, X.L., Jiang, L.B., Ye, M.X., Sun, L.D., Gragnoli, C., and Wu, R.L. (2016). Integrating evolutionary game theory into mechanistic genotype-phenotype mapping. *Trends Genet.* 32, 256–268.

Zwietering, M.H., Jongenburger, I., Rombouts, F.M., and van Riet, K. (1990). Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.* 56, 1875–1881.

## **Supplemental Information**

### **A Drive to Driven Model of Mapping**

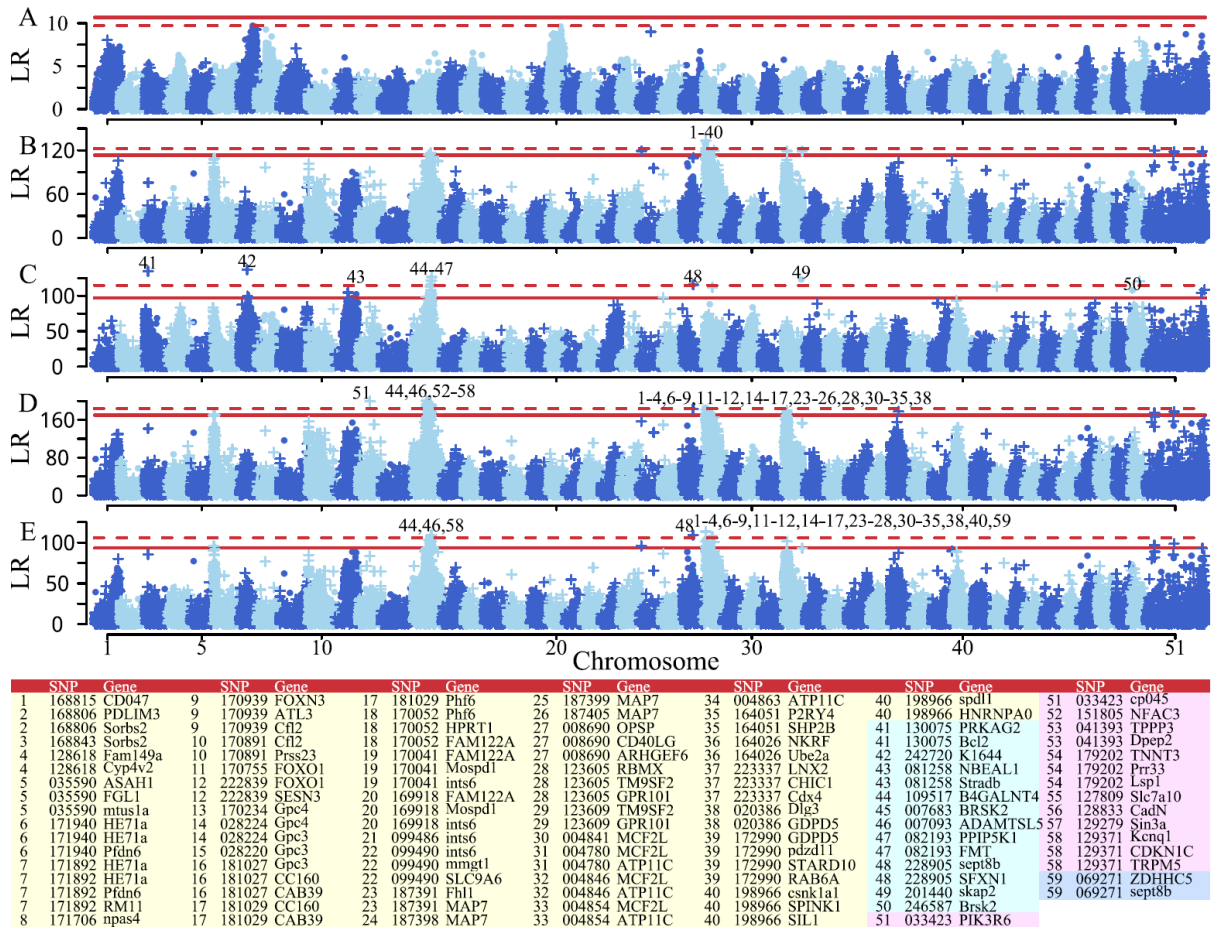
#### **Intraspecific Interaction Networks**

**Libo Jiang, Jian Xu, Mengmeng Sang, Yan Zhang, Meixia Ye, Hanyuan Zhang, Biyin Wu, Youxiu Zhu, Peng Xu, Ruyu Tai, Zixia Zhao, Yanliang Jiang, Chuanju Dong, Lidan Sun, Christopher H. Griffin, Claudia Gragnoli, and Rongling Wu**

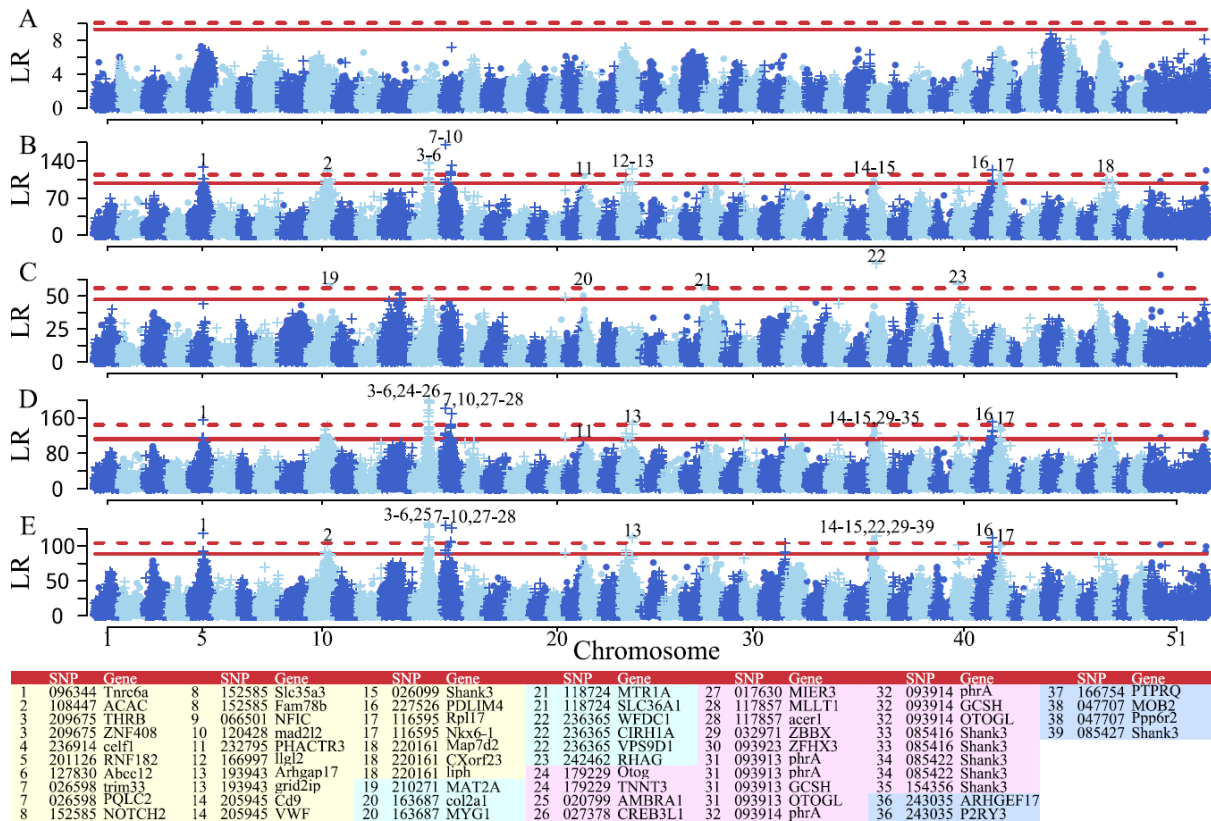
## Supplementary Figures



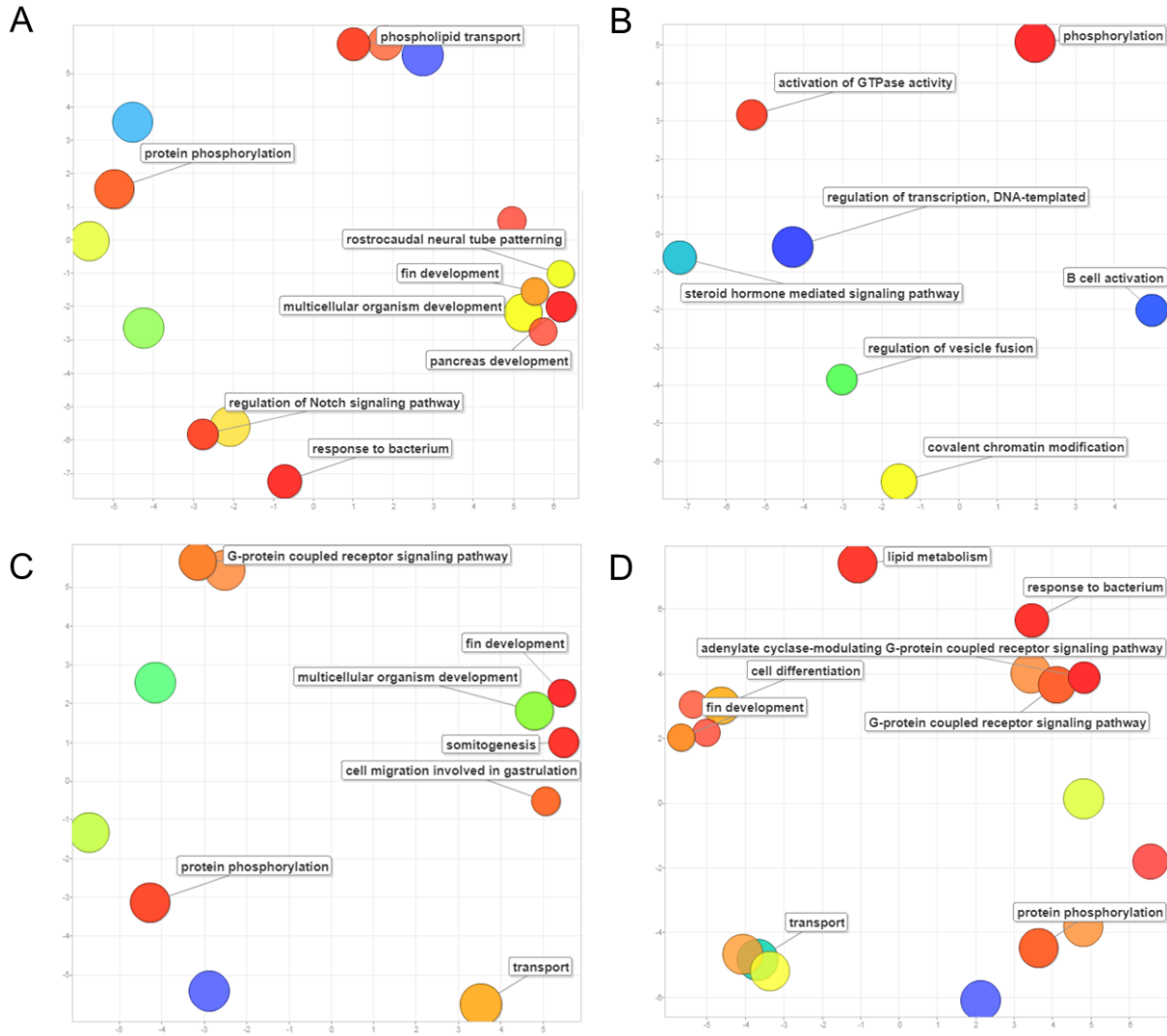
**Figure S1. Manhattan plot of log-likelihood ratios (LR) for testing significant SNPs, including testcross markers (+) and intercross markers (•), throughout the common carp genome in family H1 by the traditional mapping model (A) and our mapping model (B – E), related to Figures 4–6. Whereas the former did not discover any significant QTL, the latter has identified a number of significant loci for mutualism (B), antagonism (C), aggression (D), and altruism (E). The genome-wide critical threshold at the 5% significance level, indicated by solid lines for testcross markers and broke lines for intercross markers, was determined by 10,000 permutation tests. Through GO analysis, significant QTL for different types of social interactions, labelled by 1 – 158, are annotated by candidate genes with names given in the lower panel. Among a total of 158 QTL detected, 80 are for mutualism, 45 for antagonism, 98 for aggression and 76 for altruism, with a portion of QTL that pleiotropically affect more than one interaction type.**



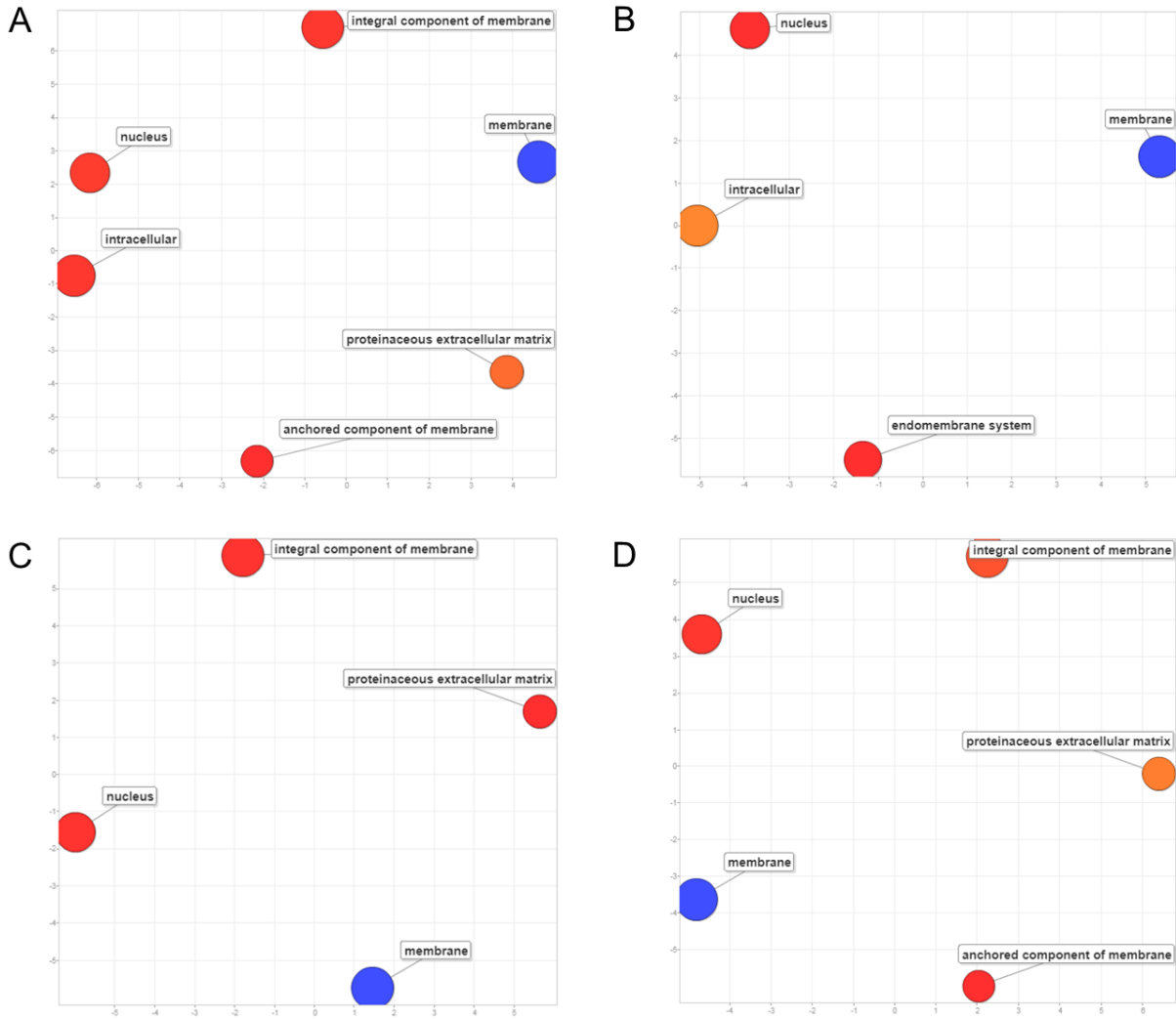
**Figure S2. Manhattan plot of log-likelihood ratios (LR) for testing significant SNPs, including testcross markers (+) and intercross markers (●) throughout the common carp genome in family G1 by the traditional mapping model (A) and our mapping model (B – E), related to Figures 4–6.** Whereas the former did not discover any significant QTL, the latter has identified a number of significant loci for mutualism (B), antagonism (C), aggression (D), and altruism (E). The genome-wide critical threshold at the 5% significance level, indicated by solid lines for testcross markers and broken lines for intercross markers, was determined by 10,000 permutation tests. Through GO analysis, significant QTL for different types of social interactions, labelled by 1 – 59, were annotated by candidate genes with names given in the lower panel. Among a total of 59 QTL detected, 40 are for mutualism, 10 for antagonism, 36 for aggression and 33 for altruism, with a portion of QTL that pleiotropically affect more than one interaction type.



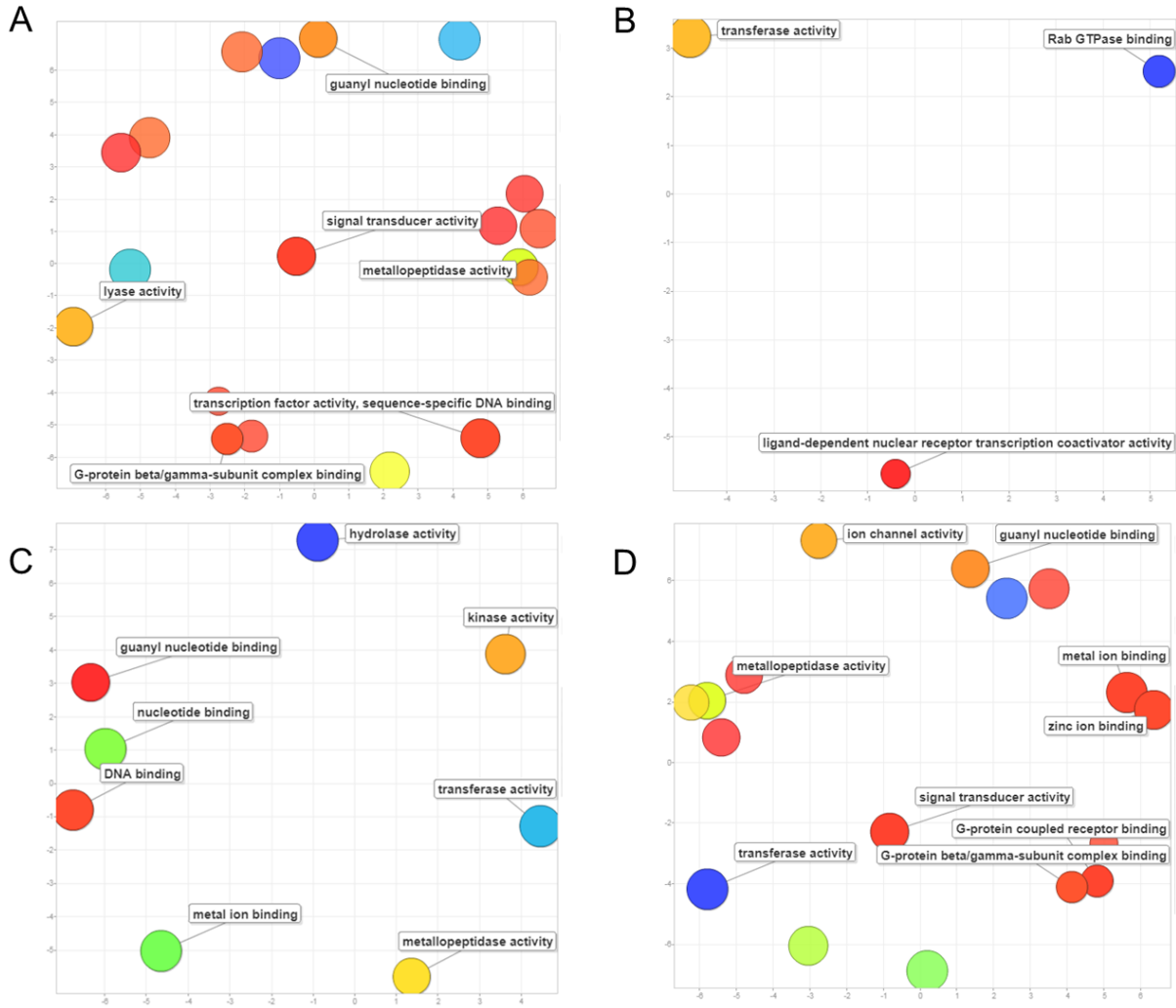
**Figure S3. Manhattan plot of log-likelihood ratios (LR) for testing significant SNPs, including testcross markers (+) and intercross markers (●) throughout the common carp genome in family Z22 by the traditional mapping model (A) and our mapping model (B – E), related to Figures 4–6.** Whereas the former did not discover any significant QTL, the latter has identified a number of significant loci for mutualism (B), antagonism (C), aggression (D), and altruism (E). The genome-wide critical threshold at the 5% significance level, indicated by solid lines for testcross markers and broke lines for intercross markers, was determined by 10,000 pertumtation tests. Through GO analysis, significant QTL for different types of social interactions, labelled by 1 – 39, were annoated to candidate genes with names given in the lower panel. Among a total of 39 QTL detected, 18 are for mutualism, 5 for antagonism, 25 for aggression and 31 for altruism, with a portion of QTL that pleiotropically affect more than one interaction type.



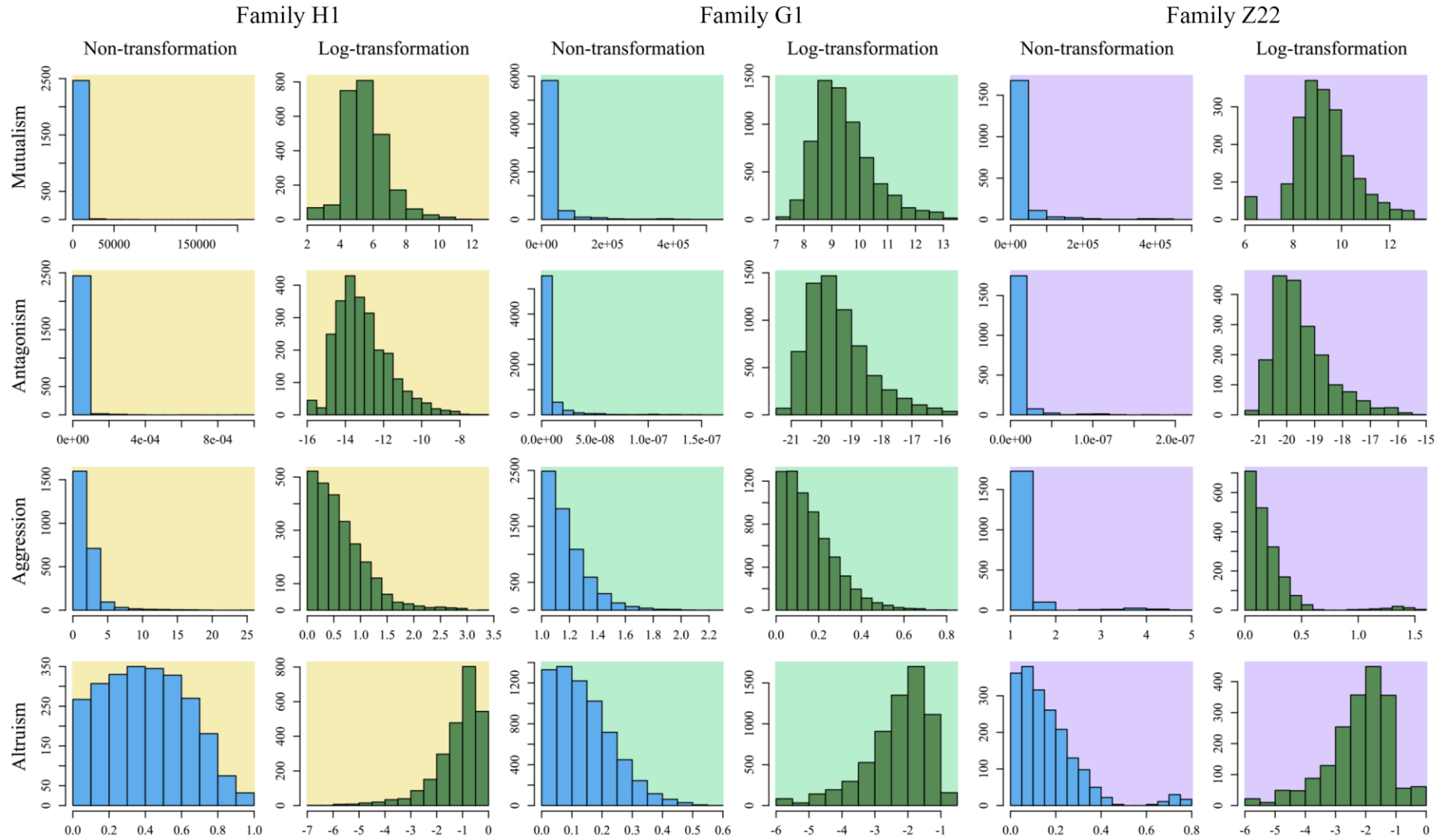
**Figure S4. Biological processes of GO terms enriched in mutualism QTL (A), antagonism QTL (B), aggression QTL (C), and altruism QTL (D) for body mass, related to Figures 4–6.** Colors of the bubbles represent the significance level of GO terms. The size of bubbles represents the gene numbers involved in each GO term. X and Y dimensions represent the semantic spaces among different GO terms.



**Figure S5. Cellular components of GO terms enriched in mutualism QTL (A), antagonism QTL (B), aggression QTL (C), and altruism QTL (D) for body mass, related to Figures 4–6.** Colors of the bubbles represent the significance level of GO terms. The size of bubbles represents the gene numbers involved in each GO term. X and Y dimensions represent the semantic spaces among different GO terms.



**Figure S6. Molecular functions of GO terms enriched in mutualism QTL (A), antagonism QTL (B), aggression QTL (C), and altruism QTL (D) for body mass, related to Figures 4–6.** Colors of the bubbles represent the significance level of GO terms. The size of bubbles represents the gene numbers involved in each GO term. X and Y dimensions represent the semantic spaces among different GO terms.



**Figure S7. Histograms of mutualistic, antagonistic, aggressive, and altruistic traits**, calculated from equation (1), and their log-transformation in three full-sib families, H1, G1, and Z22, of the common carp, related to Figures 4–6.

## Supplementary Tables

Table S1 - S9 are excel tables.

Table S10. Proportions of the total genetic variance explained by direct genetic, indirect genetic, and genome-genome (GG) epistatic genetic effects averaged over all mutualism QTL, antagonism QTL, aggression QTL, or altruism QTL for three full-sib families of the common carp, related to Figures 5 and 6.

Family	QTL type	Testcross	Intercross	Direct	Indirect	GG epistatic
H1	Mutualism	74	23	0.532±0.077	0.285±0.075	0.183±0.080
	Antagonism	42	10	0.539±0.070	0.402±0.066	0.059±0.020
	Aggression	89	25	0.515±0.079	0.252±0.075	0.233±0.105
	Altruism	70	22	0.532±0.073	0.269±0.071	0.199±0.085
G1	Mutualism	40	3	0.451±0.089	0.423±0.089	0.126±0.030
	Antagonism	4	8	0.589±0.156	0.252±0.152	0.149±0.097
	Aggression	29	13	0.468±0.085	0.389±0.110	0.143±0.044
	Altruism	30	6	0.461±0.084	0.400±0.106	0.139±0.038
Z22	Mutualism	7	13	0.572±0.158	0.307±0.118	0.121±0.092
	Antagonism	4	2	0.396±0.142	0.322±0.121	0.282±0.110
	Aggression	16	13	0.519±0.132	0.379±0.131	0.102±0.084
	Altruism	21	15	0.514±0.130	0.389±0.135	0.097±0.083

Table S11. Power comparison of QTL detection from a mapping population by a traditional model and our model under different heritability ( $H^2$ ) and sample sizes ( $m$ ), related to Figures 4–6.

	$H^2 = 0.05$		$H^2 = 0.1$	
	$m = 70$	$m = 200$	$m = 70$	$m = 200$
Social model	0.78±0.046	0.84±0.044	0.83±0.037	0.91±0.025
Traditional model	0.31±0.094	0.34±0.073	0.35±0.028	0.38±0.031

Table S12. Toy example showing how to reformat a mapping data (left panel) into the data structure of mapping social interactions (right panel). Animals in the original data (left panel) are ordered from large to small. In each pair, a larger animal is arrayed in column L and a smaller one in column S (right panel), related to Figures 1–3. See Fig. 1 for the definition of different types of interactions.

Progeny	Marker	Phenotype	Social Interaction							Pair with L and S	
			No.	Pair	G×G	Mutualism ( $z_{mu}$ )	Antagonism ( $z_{an}$ )	Aggression ( $z_{ag}$ )	Altruism ( $z_{mu}$ )	x\L	y\S
1	AA	$w_1$	1	1×2	$AA \times Aa$	$w_1 w_2 / (w_1 - w_2)$	$1 / (w_1 w_2 (w_1 - w_2))$	$w_1 / w_2$	$1 - w_2 / w_1$	$w_1$	$w_2$
2	Aa	$w_2$	2	1×3	$AA \times Aa$	$w_1 w_3 / (w_1 - w_3)$	$1 / (w_1 w_3 (w_1 - w_3))$	$w_1 / w_3$	$1 - w_3 / w_1$	$w_1$	$w_3$
3	Aa	$w_3$	3	1×4	$AA \times AA$	$w_1 w_4 / (w_1 - w_4)$	$1 / (w_1 w_4 (w_1 - w_4))$	$w_1 / w_4$	$1 - w_4 / w_1$	$w_1$	$w_4$
4	AA	$w_4$	4	1×5	$AA \times Aa$	$w_1 w_5 / (w_1 - w_5)$	$1 / (w_1 w_5 (w_1 - w_5))$	$w_1 / w_5$	$1 - w_5 / w_1$	$w_1$	$w_5$
5	Aa	$w_5$	5	1×6	$AA \times AA$	$w_1 w_6 / (w_1 - w_6)$	$1 / (w_1 w_6 (w_1 - w_6))$	$w_1 / w_6$	$1 - w_6 / w_1$	$w_1$	$w_6$
6	AA	$w_6$	6	2×3	$Aa \times Aa$	$w_2 w_3 / (w_2 - w_3)$	$1 / (w_2 w_3 (w_2 - w_3))$	$w_2 / w_3$	$1 - w_3 / w_2$	$w_2$	$w_3$
			7	2×4	$Aa \times AA$	$w_2 w_4 / (w_2 - w_4)$	$1 / (w_2 w_4 (w_2 - w_4))$	$w_2 / w_4$	$1 - w_4 / w_2$	$w_2$	$w_4$
			8	2×5	$Aa \times Aa$	$w_2 w_5 / (w_2 - w_5)$	$1 / (w_2 w_5 (w_2 - w_5))$	$w_2 / w_5$	$1 - w_5 / w_2$	$w_2$	$w_5$
			9	2×6	$Aa \times AA$	$w_2 w_6 / (w_2 - w_6)$	$1 / (w_2 w_6 (w_2 - w_6))$	$w_2 / w_6$	$1 - w_6 / w_2$	$w_2$	$w_6$
			10	3×4	$Aa \times AA$	$w_3 w_4 / (w_3 - w_4)$	$1 / (w_3 w_4 (w_3 - w_4))$	$w_3 / w_4$	$1 - w_4 / w_3$	$w_3$	$w_4$
			11	3×5	$Aa \times Aa$	$w_3 w_5 / (w_3 - w_5)$	$1 / (w_3 w_5 (w_3 - w_5))$	$w_3 / w_5$	$1 - w_5 / w_3$	$w_3$	$w_5$
			12	3×6	$Aa \times AA$	$w_3 w_6 / (w_3 - w_6)$	$1 / (w_3 w_6 (w_3 - w_6))$	$w_3 / w_6$	$1 - w_6 / w_3$	$w_3$	$w_6$
			13	4×5	$AA \times Aa$	$w_4 w_5 / (w_4 - w_5)$	$1 / (w_4 w_5 (w_4 - w_5))$	$w_4 / w_5$	$1 - w_5 / w_4$	$w_4$	$w_5$
			14	4×6	$AA \times AA$	$w_4 w_6 / (w_4 - w_6)$	$1 / (w_4 w_6 (w_4 - w_6))$	$w_4 / w_6$	$1 - w_6 / w_4$	$w_4$	$w_6$
			15	5×6	$Aa \times Aa$	$w_5 w_6 / (w_5 - w_6)$	$1 / (w_5 w_6 (w_5 - w_6))$	$w_5 / w_6$	$1 - w_6 / w_5$	$w_5$	$w_6$

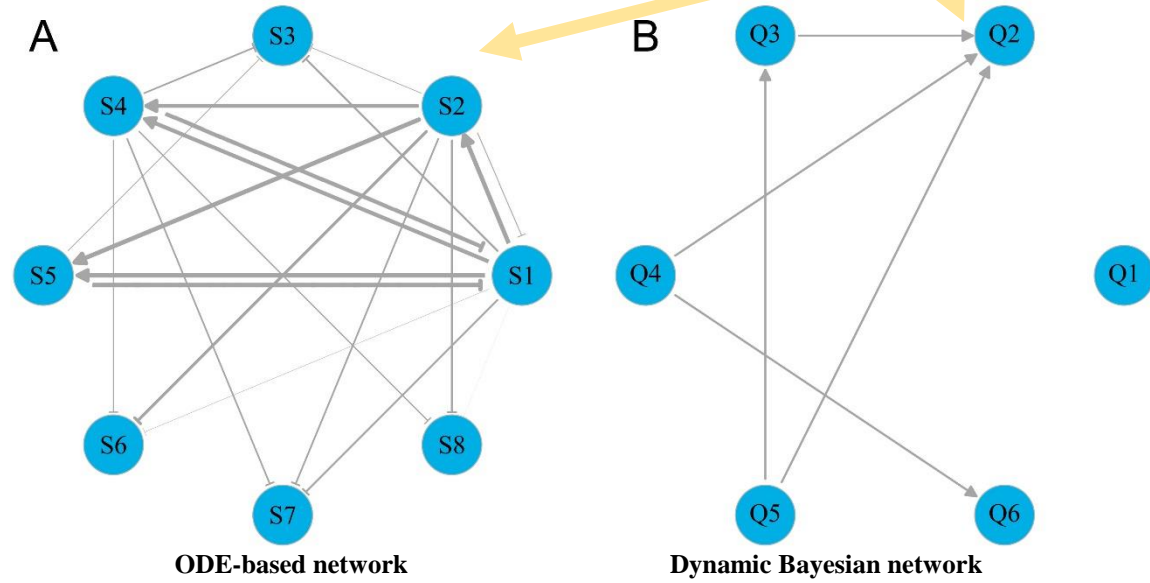
Table S13. Genotypic values of four GG combinations and their underlying components: direct genetic, indirect genetic, and genome-genome (gg) epistatic effects, derived from animal pairs, each with a larger one arrayed in column L and a smaller one in column S, related to Figures 1–3.

GG Combination	Column L	Column S
<b>Testcross Marker</b>		
AA × AA	$\mu_{11}^x = \mu_x + a_{x \leftarrow x} + a_{x \leftarrow y} + e_{aa}^x$	$\mu_{11}^y = \mu_y + a_{y \leftarrow x} + a_{y \leftarrow y} + e_{aa}^y$
AA × Aa	$\mu_{12}^x = \mu_x + a_{x \leftarrow x} - a_{x \leftarrow y} - e_{aa}^x$	$\mu_{12}^y = \mu_y + a_{y \leftarrow x} - a_{y \leftarrow y} - e_{aa}^y$
Aa × AA	$\mu_{21}^x = \mu_x - a_{x \leftarrow x} + a_{x \leftarrow y} - e_{aa}^x$	$\mu_{21}^y = \mu_y - a_{y \leftarrow x} + a_{y \leftarrow y} - e_{aa}^y$
Aa × Aa	$\mu_{22}^x = \mu_x - a_{x \leftarrow x} - a_{x \leftarrow y} + e_{aa}^x$	$\mu_{22}^y = \mu_y - a_{y \leftarrow x} - a_{y \leftarrow y} + e_{aa}^y$
Overall Mean	$\mu_x = \frac{1}{4}(\mu_{11}^x + \mu_{12}^x + \mu_{21}^x + \mu_{22}^x)$	$\mu_y = \frac{1}{4}(\mu_{11}^y + \mu_{12}^y + \mu_{21}^y + \mu_{22}^y)$
Direct Effect	$a_{x \leftarrow x} = \frac{1}{4}(\mu_{11}^x + \mu_{12}^x - \mu_{21}^x - \mu_{22}^x)$	$a_{y \leftarrow y} = \frac{1}{4}(\mu_{11}^y + \mu_{12}^y - \mu_{21}^y - \mu_{22}^y)$
Indirect Effect	$a_{x \leftarrow y} = \frac{1}{4}(\mu_{11}^x - \mu_{12}^x + \mu_{21}^x - \mu_{22}^x)$	$a_{y \leftarrow x} = \frac{1}{4}(\mu_{11}^y + \mu_{12}^y - \mu_{21}^y - \mu_{22}^y)$
gg Epistasis	$e_{aa}^x = \frac{1}{4}(\mu_{11}^x - \mu_{12}^x - \mu_{21}^x + \mu_{22}^x)$	$e_{aa}^y = \frac{1}{4}(\mu_{11}^y - \mu_{12}^y - \mu_{21}^y + \mu_{22}^y)$
<b>Intercross Marker</b>		
AA × AA	$\mu_{11}^x = \mu_x + a_{x \leftarrow x} + a_{x \leftarrow x} + e_{aa}^x$	$\mu_{11}^y = \mu_y + a_{y \leftarrow x} + a_{y \leftarrow y} + e_{aa}^y$
AA × Aa	$\mu_{12}^x = \mu_x + a_{x \leftarrow x} + d_{x \leftarrow y} + e_{ad}^x$	$\mu_{12}^y = \mu_y + a_{y \leftarrow x} + d_{y \leftarrow y} + e_{ad}^y$
AA × aa	$\mu_{13}^x = \mu_x + a_{x \leftarrow x} - a_{x \leftarrow y} - e_{aa}^x$	$\mu_{13}^y = \mu_y + a_{y \leftarrow x} - a_{y \leftarrow x} - e_{aa}^x$
Aa × AA	$\mu_{21}^x = \mu_x + d_{x \leftarrow x} + a_{x \leftarrow y} + e_{da}^x$	$\mu_{21}^y = \mu_y + d_{y \leftarrow x} + a_{y \leftarrow y} + e_{da}^y$
Aa × Aa	$\mu_{22}^x = \mu_x + d_{x \leftarrow x} + d_{x \leftarrow y} + e_{dd}^x$	$\mu_{22}^y = \mu_y + d_{y \leftarrow x} + d_{y \leftarrow y} + e_{dd}^y$
Aa × aa	$\mu_{23}^x = \mu_x + d_{x \leftarrow x} - a_{x \leftarrow y} - e_{da}^x$	$\mu_{23}^y = \mu_y + d_{y \leftarrow x} - a_{y \leftarrow y} + e_{da}^y$
aa × AA	$\mu_{31}^x = \mu_x - a_{x \leftarrow x} + a_{x \leftarrow y} - e_{aa}^x$	$\mu_{31}^y = \mu_y - a_{y \leftarrow x} + a_{y \leftarrow y} - e_{aa}^y$
aa × Aa	$\mu_{32}^x = \mu_x - a_{x \leftarrow x} + d_{x \leftarrow y} - e_{ad}^x$	$\mu_{32}^y = \mu_y - a_{y \leftarrow x} + d_{y \leftarrow y} - e_{ad}^y$
aa × aa	$\mu_{33}^x = \mu_x - a_{x \leftarrow x} - a_{x \leftarrow y} + e_{aa}^x$	$\mu_{33}^y = \mu_y - a_{y \leftarrow x} - a_{y \leftarrow y} + e_{aa}^y$
Overall Mean	$\mu_x = \frac{1}{9} \sum_{j_1=1}^3 \sum_{j_2=1}^3 \mu_{j_1 j_2}^x$	$\mu_y = \frac{1}{9} \sum_{j_1=1}^3 \sum_{j_2=1}^3 \mu_{j_1 j_2}^y$
Direct Effect	$a_{x \leftarrow x} = \frac{1}{4}(\mu_{11}^x + \mu_{13}^x - \mu_{31}^x - \mu_{33}^x)$ $d_{x \leftarrow y} = \frac{1}{4}[2(\mu_{12}^x - \mu_{32}^x) - (\mu_{11}^x + \mu_{13}^x - \mu_{31}^x - \mu_{33}^x)]$	$a_{y \leftarrow y} = \frac{1}{4}(\mu_{11}^y + \mu_{13}^y - \mu_{31}^y - \mu_{33}^y)$ $d_{y \leftarrow x} = \frac{1}{4}[2(\mu_{12}^y - \mu_{32}^y) - (\mu_{11}^y + \mu_{13}^y - \mu_{31}^y - \mu_{33}^y)]$
Indirect Effect	$a_{x \leftarrow y} = \frac{1}{4}(\mu_{11}^x + \mu_{13}^x - \mu_{31}^x - \mu_{33}^x)$ $d_{x \leftarrow y} = \frac{1}{4}[2(\mu_{21}^x - \mu_{23}^x) - (\mu_{11}^x + \mu_{31}^x - \mu_{13}^x - \mu_{33}^x)]$	$a_{y \leftarrow x} = \frac{1}{4}(\mu_{11}^y + \mu_{13}^y - \mu_{31}^y - \mu_{33}^y)$ $d_{y \leftarrow x} = \frac{1}{4}[2(\mu_{21}^y - \mu_{23}^y) - (\mu_{11}^y + \mu_{31}^y - \mu_{13}^y - \mu_{33}^y)]$
gg Epistasis	$e_{aa}^x = \frac{1}{4}(\mu_{11}^x + \mu_{33}^x - \mu_{13}^x - \mu_{31}^x)$ $e_{ad}^x = \frac{1}{4}[2(\mu_{12}^x - \mu_{32}^x) - (\mu_{11}^x + \mu_{13}^x - \mu_{31}^x - \mu_{33}^x)]$ $e_{da}^x = \frac{1}{4}[2(\mu_{21}^x - \mu_{23}^x) - (\mu_{31}^x + \mu_{33}^x - \mu_{11}^x - \mu_{13}^x)]$ $e_{dd}^x = \frac{1}{4}(2\mu_{22}^x - \mu_{12}^x - \mu_{21}^x - \mu_{23}^x - \mu_{32}^x)$	$e_{aa}^y = \frac{1}{4}(\mu_{11}^y + \mu_{33}^y - \mu_{13}^y - \mu_{31}^y)$ $e_{ad}^y = \frac{1}{4}[2(\mu_{12}^y - \mu_{32}^y) - (\mu_{11}^y + \mu_{13}^y - \mu_{31}^y - \mu_{33}^y)]$ $e_{da}^y = \frac{1}{4}[2(\mu_{21}^y - \mu_{23}^y) - (\mu_{31}^y + \mu_{33}^y - \mu_{11}^y - \mu_{13}^y)]$ $e_{dd}^y = \frac{1}{4}(2\mu_{22}^y - \mu_{12}^y - \mu_{21}^y - \mu_{23}^y - \mu_{32}^y)$

Note: for a testcross QTL,  $\mu_{j_1 j_2}^x$  and  $\mu_{j_1 j_2}^y$  ( $j_1, j_2 = 1, 3$ ) are the genotypic values of GG combinations for column A and B, respectively;  $a_{x \leftarrow x}$  and  $a_{y \leftarrow y}$  are the direct effects of columns A and B on their own phenotype;  $a_{x \leftarrow y}$  and  $a_{y \leftarrow x}$  are the indirect effects of columns B and A on the phenotype of its counterpart; and  $e_{aa}^x$  and  $e_{aa}^y$  are the genome-genome epistatic effects due to the interactions between columns A's and B's alleles on the phenotype of columns A and B, respectively. All of these definitions can be extended to an intercross QTL that includes additive ( $a$ ), dominant ( $d$ ), genome-genome additive × additive ( $e_{aa}$ ), genome-genome additive × dominant ( $e_{ad}$ ), genome-genome dominant × additive ( $e_{da}$ ), and genome-genome dominant × dominant ( $e_{dd}$ ) effects.

Table S14. Data structure of a toy mapping population used to infer a directed acyclic graph, related to Figures 5 and 6. Social network (**A**) was constructed from ODEs, whereas QTL networks (**B**) was constructed from dynamic Bayesian networks.  $\sigma_Q^2$  is the genetic variance of a locus and  $\mu$  is the population mean.

No.	QTL						Phenotype	QTL					
	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$		$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_6$
$S_1$	AA	Aa	AA	AA	aa	AA	1.5	1.70	1.75	1.70	1.81	1.70	1.77
$S_2$	aa	AA	aa	AA	Aa	aa	2.12	1.94	1.94	1.86	1.81	1.88	1.98
$S_3$	Aa	Aa	Aa	Aa	Aa	Aa	1.85	1.89	1.75	1.87	1.86	1.88	1.81
$S_4$	Aa	aa	Aa	aa	AA	AA	1.91	1.89	1.84	1.87	1.81	1.88	1.77
$S_5$	aa	AA	aa	aa	Aa	Aa	1.76	1.94	1.94	1.86	1.81	1.88	1.81
$S_6$	AA	aa	Aa	Aa	AA	aa	1.84	1.70	1.84	1.87	1.86	1.88	1.98
$S_7$	Aa	Aa	AA	Aa	aa	AA	1.9	1.89	1.75	1.70	1.86	1.70	1.77
$S_8$	AA	aa	aa	aa	Aa	AA	1.77	1.70	1.84	1.86	1.81	1.88	1.77
$\sigma_Q^2 = 0.011$		0.006	0.007	0.001	0.007	0.009	$\mu = 1.83$	$\sigma_Q = 0.108$	0.079	0.081	0.026	0.082	0.093



## Transparent Methods

### Mapping design

Consider a full-sib mapping population of animals genotyped for SNPs throughout the whole genome. By rearing them in a common environment, these animals are allowed to randomly interact with each other. The animals' resource- and space-responsive phenotypes, such as body weight and body length, are measured after the population has experienced full mutual interactions during ontogeny. To illustrate our new mapping theory, we designed a toy example in which six animals are genotyped by a testcross marker with two genotypes  $AA$  and  $Aa$  and phenotyped by a growth-related trait. Without loss of generality, we labelled and arranged the six animals in order from high to small phenotypes denoted as  $w_1, \dots, w_6$  (left, **Table S12**). By pairing all animals, we reformatted genotype and phenotype data across a total of  $(5 \times 6)/2 = 15$  pairs, where a larger animal in each pair is arrayed in left column and a smaller one in right column (right, **Table S12**). Let  $w_L$  and  $w_S$  denote the phenotypic values of the larger animal  $L$  ( $L = 1, \dots, 5$ ) and smaller animal  $S$  ( $S = 2, \dots, 6$ ) from a pair, respectively. For a particular pair, we use the mathematic expressions, as shown in Fig. 1, to calculate the parameters  $z_{mu}$ ,  $z_{an}$ ,  $z_{ag}$ , and  $z_{al}$  that describe the strengths of their mutualism, antagonism, aggression, and altruism, respectively. Among all pairs, two genotypes at the marker form four possible genotype ( $G$ )  $\times$  genotype ( $G$ ) combinations,  $AA \times AA$  (coded as  $1 \times 1$ ),  $AA \times Aa$  (coded as  $1 \times 2$ ),  $Aa \times AA$  (coded as  $2 \times 1$ ), and  $Aa \times Aa$  (coded as  $2 \times 2$ ). A traditional mapping approach is to associate marker genotypes with trait phenotypes across individual animals (left, **Table S12**), whereas our new mapping model performs the analysis of association between GG combinations and the derived phenotypes of mutualism, antagonism, aggression, and altruism across animal-animal pairs (right, **Table S12**).

### Statistical mapping of social interactions

**Likelihood:** We use  $z_i$  to denote the value of a derivative trait (i.e., mutualism, antagonism, aggression, or altruism) for pair  $i$  ( $i = 1, \dots, n$ ). Let  $n_{1 \times 1}$ ,  $n_{1 \times 2}$ ,  $n_{2 \times 1}$ , and  $n_{2 \times 2}$  denote the observations of GG combinations,  $AA \times AA$ ,  $AA \times Aa$ ,  $Aa \times AA$ , and  $Aa \times Aa$  at a test marker, respectively. The likelihood of a derivative trait at this marker is formulated as

$$L(z) = \sum_{i=1}^{n_{1 \times 1}} f_{1 \times 1}(z_i) \sum_{i=1}^{n_{1 \times 2}} f_{1 \times 2}(z_i) \sum_{i=1}^{n_{2 \times 1}} f_{2 \times 1}(z_i) \sum_{i=1}^{n_{2 \times 2}} f_{2 \times 2}(z_i) \quad (1)$$

where  $f(z_i)$  is the probability density of the derivative trait for a particular GG combination. The four derivative traits may have complicated forms of density function. However, we first test if they are normally distributed after log-transformation and implement the normal density function if they pass the test. In the subsequent data analysis of carp fish body mass, we find that all derivative traits, except for aggression, approximately follow a normal distribution after they are log-transformed (Fig. S7). The likelihood (1) has four genotypic values of GG combinations for a derivative trait, denoted as  $\mu_{1 \times 1}^z$ ,  $\mu_{1 \times 2}^z$ ,  $\mu_{2 \times 1}^z$ , and  $\mu_{2 \times 2}^z$ , respectively, and a residual variance. A standard likelihood approach is implemented to obtain the maximum likelihood estimates (MLEs) of these parameters.

As a ratio trait between two variables, aggression can be approached by the Cauchy density function. Let  $x$  and  $y$  denote the trait values of animal column  $L$  and  $S$ , respectively. For ratio  $z_i = y_i/x_i$  ( $x_i > y_i$ ), the density function of GG combination  $j_1 j_2$ ,  $f_{j_1 j_2}(z_i)$ , is a product of a Cauchy density and a complicated function (Cedilnik et al., 2004), expressed as

$$\begin{aligned}
f_{j_1 \times j_2}(z_i) &= \frac{\sigma_x \sigma_y \sqrt{1 - \rho^2}}{\pi(\sigma_x^2 z_i^2 - 2\rho\sigma_x\sigma_y z_i + \sigma_y^2)} \left[ \exp\left(-\frac{1}{2} \sup R_{j_1 \times j_2}^2\right) \left(1 + \frac{R_{j_1 \times j_2} \Phi(R_{j_1 \times j_2})}{\phi(R_{j_1 \times j_2})}\right) \right] \\
&= \frac{\sigma_x \sigma_y \sqrt{1 - \rho^2}}{\pi(\sigma_y^2 z_i^2 - 2\rho\sigma_x\sigma_y z_i + \sigma_x^2)} \left[ \exp\left(-\frac{1}{2} \sup R_{j_1 \times j_2}^2\right) \right. \\
&\quad \left. + \sqrt{2\pi} R_{j_1 \times j_2} \Phi(R_{j_1 \times j_2}) \exp\left(-\frac{1}{2} [\sup R_{j_1 \times j_2}^2 - R_{j_1 \times j_2}^2]\right) \right] \quad (2)
\end{aligned}$$

where

$$\begin{aligned}
R_{j_1 \times j_2} &= \frac{(\sigma_y^2 \mu_{j_1 \times j_2}^x - \rho\sigma_x\sigma_y \mu_{j_1 \times j_2}^y) z_i - \rho\sigma_x\sigma_y \mu_{j_1 \times j_2}^x + \sigma_x^2 \mu_{j_1 \times j_2}^y}{\sigma_x \sigma_y \sqrt{1 - \rho^2} \sqrt{\sigma_y^2 z_i^2 - 2\rho\sigma_x\sigma_y z_i + \sigma_x^2}} \\
&= \frac{\left(\frac{\mu_{j_1 \times j_2}^x}{\sigma_x} - \rho \frac{\mu_{j_1 \times j_2}^y}{\sigma_y}\right) z_i - \left(\rho \frac{\mu_{j_1 \times j_2}^x}{\sigma_x} - \frac{\mu_{j_1 \times j_2}^y}{\sigma_y}\right) \frac{\sigma_x}{\sigma_y}}{\sqrt{1 - \rho^2} \sqrt{z_i^2 - 2\rho \frac{\sigma_x}{\sigma_y} z_i + \left(\frac{\sigma_x}{\sigma_y}\right)^2}}
\end{aligned}$$

$$\begin{aligned}
\sup R_{j_1 \times j_2}^2 &= \frac{\sigma_y^2 \mu_{j_1 \times j_2}^x - 2\rho\sigma_x\sigma_y \mu_{j_1 \times j_2}^x \mu_{j_1 \times j_2}^y + \sigma_x^2 \mu_{j_1 \times j_2}^y}{\sigma_x^2 \sigma_y^2 (1 - \rho)} \\
&= \frac{\left(\frac{\mu_{j_1 \times j_2}^x}{\sigma_x}\right)^2 - 2 \frac{\mu_{j_1 \times j_2}^x}{\sigma_x} \frac{\mu_{j_1 \times j_2}^y}{\sigma_y} + \left(\frac{\mu_{j_1 \times j_2}^y}{\sigma_y}\right)^2}{1 - \rho^2}
\end{aligned}$$

$$\sup R_{j_1 \times j_2}^2 - R_{j_1 \times j_2}^2 = \frac{(\mu_{j_1 \times j_2}^y - \mu_{j_1 \times j_2}^x z_i)^2}{\sigma_x^2 z_i^2 - 2\rho\sigma_y\sigma_x z_i + \sigma_y^2} = \frac{\left(\frac{\mu_{j_1 \times j_2}^y}{\sigma_y} \frac{\sigma_y}{\sigma_x} - \frac{\mu_{j_1 \times j_2}^x}{\sigma_x} z_i\right)^2}{z_i^2 - 2\rho \frac{\sigma_y}{\sigma_x} z_i + \left(\frac{\sigma_y}{\sigma_x}\right)^2},$$

and

$$\Phi(R_{j_1 \times j_2}) = \int_0^{R_{j_1 \times j_2}} \phi(r) dr = \int_0^{R_{j_1 \times j_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} dr = \frac{1}{2} \operatorname{erf}\left(\frac{R_{j_1 \times j_2}}{\sqrt{2}}\right)$$

is the error function. The density function (2) was implemented into the likelihood (1) that is defined by GG combination-dependent genotypic means of variables  $x$  ( $\mu_{j_1 \times j_2}^x$ ) and  $y$  ( $\mu_{j_1 \times j_2}^y$ ), the residual variances of variables  $x$  ( $\sigma_x^2$ ) and  $y$  ( $\sigma_y^2$ ), and the correlation between the two variables ( $\rho$ ).

**Significance test:** To determine whether a significant QTL exists to affect a type of social interaction, we compare the genotypic difference among GG combinations. For mutualism, antagonism, and altruism traits, four genotypic values of the derivative trait are estimated directly. For the aggression trait, its genotypic values are estimated as  $\mu_{j_1 \times j_2}^z = \mu_{j_1 \times j_2}^y / \mu_{j_1 \times j_2}^x$ . In general, we formulate the following hypotheses for significance test:

$$\begin{aligned}
H_0: \mu_{1 \times 1}^z &= \mu_{1 \times 2}^z = \mu_{2 \times 1}^z = \mu_{2 \times 2}^z = \mu^z \\
H_1: &\text{At least one of the above equalities does not hold.} \quad (3)
\end{aligned}$$

A log-likelihood ratio calculated from the  $H_0$  (there is no QTL) and  $H_1$  hypothesis (there is a QTL) is used to test if these GG combinations differ from each other for a derivative trait. If the null hypothesis is rejected, then we could claim the existence of a significant QTL that

affects the derivative trait. We call such a QTL a mutualism QTL, antagonism QTL, aggression QTL, or altruism QTL if the derivative trait is  $z_{mu}$ ,  $z_{an}$ ,  $z_{ag}$ , or  $z_{al}$ , respectively. The critical threshold for the significance test can be empirically determined through permutation tests.

### Quantitative genetic dissection of social interactions

After a significant QTL for an interaction parameter is detected, the new theory can be used to test how this QTL affects phenotypic variation. As described above, each pair is composed of two animals, a larger one L arrayed in left column (with trait value denoted as  $x$ ) and a smaller one S in right column (with trait value denoted as  $y$ ) (right, **Table S12**). For a QTL significant by test, we calculate the MLEs of the genotypic value of each GG combination for each column, i.e.,  $\mu_{j_1 j_2}^x$  for column L and  $\mu_{j_1 j_2}^y$  for column S ( $j_1, j_2 = 1$  for AA, 2 for Aa). According to quantitative genetic theory, we partitioned these genotypic values into their underlying components (**Table S13**), including the overall means for different columns, denoted as  $\mu_x$  for column L and  $\mu_y$  for column S; direct genetic effects of QTL alleles from two columns on their own phenotypes, denoted as  $a_{x \leftarrow x}$  for column L and  $a_{y \leftarrow y}$  for column S; indirect genetic effects of QTL alleles from two columns on each other's phenotypes, denoted as  $a_{x \leftarrow y}$  for column S affecting column L and  $a_{y \leftarrow x}$  for column L affecting column S; and genome-genome epistatic effects of QTL alleles from different columns, denoted as  $e_{aa}^x$  on column L and  $e_{aa}^y$  on column S.

Based on the component structure of a GG combination genotypic value (**Table S13**), we solve these effect parameters by

$$\begin{bmatrix} a_{x \leftarrow x} \\ a_{x \leftarrow y} \\ e_{aa}^x \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11}^x \\ \mu_{12}^x \\ \mu_{21}^x \\ \mu_{22}^x \end{bmatrix}, \begin{bmatrix} a_{y \leftarrow x} \\ a_{y \leftarrow y} \\ e_{aa}^y \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11}^y \\ \mu_{12}^y \\ \mu_{21}^y \\ \mu_{22}^y \end{bmatrix} \quad (4)$$

After these effect parameters are estimated, we formulate a procedure to test the significance of each of them by a log-likelihood ratio approach. For example, the null hypotheses for testing these effects are expressed as

$$H_0: a_{x \leftarrow x} = a_{y \leftarrow y} = 0, \text{ for the direct effect} \quad (5)$$

$$H_0: a_{x \leftarrow y} = a_{y \leftarrow x} = 0, \text{ for the indirect effect} \quad (6)$$

$$H_0: e_{aa}^x = e_{aa}^y = 0, \text{ for the fish-fish epistatic effect} \quad (7)$$

The critical thresholds for all the above hypotheses tests can be obtained from classic chi-square statistics or simulation studies. If these effects are significant, we calculate their means over two columns, i.e.,  $a_D = (a_{x \leftarrow x} + a_{y \leftarrow y})/2$ ,  $a_I = (a_{x \leftarrow y} + a_{y \leftarrow x})/2$ , and  $e_{aa} = (e_{aa}^x + e_{aa}^y)/2$ , as the estimates of direct, indirect, and genome-genome epistatic effects on a phenotypic trait in the population.

We next describe a procedure to estimate genetic variances due to direct, indirect, and genome-genome epistatic effects. Considering columns L and S of the right part of **Table S12**, we calculate the genetic variance among the four GG combinations at a significant QTL, denoted as  $V_G^x$  for column L and  $V_G^y$  for column S. The mean of  $V_G^x$  and  $V_G^y$ , expressed as  $V_G$ , is the estimation of the total genetic variance explained by the QTL. Based on column L's and S's genotypes, we calculate the genetic variance of the trait in columns L and S, respectively, denoted as  $V_{x \leftarrow x}$  for column L and  $V_{y \leftarrow y}$  for column S, whose mean is the estimated direct genetic variance of the QTL, expressed as  $V_D$ . Similarly, using column L's and S's genotypes, we can calculate the genetic variance of the trait in columns S and L, respectively, denoted as  $V_{y \leftarrow x}$  for column L affecting S and  $V_{x \leftarrow y}$  for column S affecting L. The mean of these two genetic

variances, expressed as  $V_I$ , is the estimated indirect genetic variance of the QTL. For columns L and S, we calculate  $V_{aa}^x = V_G^x - V_{x \leftarrow x} - V_{x \leftarrow y}$  and  $V_{aa}^y = V_G^y - V_{y \leftarrow y} - V_{y \leftarrow x}$ , respectively, and their mean is the genome-genome epistatic genetic variance of the QTL, denoted as  $V_{aa}$ . We can further calculate the proportions of direct, indirect, and genome-genome epistatic effects to the total genetic variance by this QTL.

For an outcrossing species like the carp fish, a full-sib family population derived from two heterozygous parents may include two types of markers, i.e., testcross markers at which one parent is heterozygous whereas the other is homozygous, and intercross markers at which both parents are heterozygous (Wu et al., 2002; Lu et al., 2004). The procedure described above can be similarly used to map mutualism, antagonism, aggression, or altruism QTL based on intercross markers. For an intercross marker with three genotypes (AA, Aa, and aa), we use nine GG combinations, AA  $\times$  AA (coded as 1 $\times$ 1), AA  $\times$  Aa (coded as 1 $\times$ 2), AA  $\times$  aa (coded as 1 $\times$ 3), Aa  $\times$  AA (coded as 2 $\times$ 1), Aa  $\times$  Aa (coded as 2 $\times$ 2), Aa  $\times$  aa (coded as 2 $\times$ 3), aa  $\times$  Aa (coded as 3 $\times$ 1), aa  $\times$  Aa (coded as 3 $\times$ 2), and aa  $\times$  aa (coded as 3 $\times$ 3). Similarly, we formulated a log-likelihood approach to estimate the MLEs of genotypic values of nine GG combinations for columns L and S, and tested and estimated the significance of direct additive ( $a_D$ ) and dominant genetic effects ( $d_D$ ), indirect additive ( $a_I$ ), and dominant genetic effects ( $d_I$ ), and genome-genome additive-additive ( $e_{AA}$ ), genome-genome additive-dominant ( $e_{AD}$ ), genome-genome dominant-additive ( $e_{DA}$ ), and genome-genome dominant-dominant epistatic genetic effects ( $e_{DD}$ ) (**Table S13**).

### Inferring directed acyclic networks

To better understand how different types of QTL, mutualistic, antagonistic, aggressive, or altruistic, jointly affect the phenotypic trait of animals, we develop and implement a statistical algorithm to infer a directed acyclic graph (DAG) of QTL interactions. We use a toy example to explain our algorithm. Suppose there are six QTL under consideration, each with three genotypes AA (coded as 1), Aa (coded as 2) and aa (coded as 3). These QTL are segregating in the mapping population of eight phenotyped animals (left, **Table S14**). For the trait measured, we calculate its population mean ( $\mu$ ) averaged over all animals and also its marginal genotypic means  $\mu_{jk}$  ( $j = 1, 2, 3$ ) over the animals carrying the same genotype at each QTL  $k$ . Now, we assign each genotype at each QTL by its marginal genotypic mean to form an (8  $\times$  6) matrix of genotypic values (right, **Table S14**), from which two types of DAG, constructed by different QTL and different animals, respectively, were inferred by Bayesian networks.

**QTL network:** The structure of a QTL network is defined by two sets: the set of nodes (vertices) represented by individual QTL and the set of directed edges of dependence (directed epistasis) among the QTL. Because each QTL has three distinct genotypes, its marginal genotypic values (right, **Table S14**) can be better viewed as ordinal variables. The most general approach for constructing ordinal Bayesian networks is to treat ordinal variables as nominal so that nominal techniques can be used. However, this treatment entails a loss of information because the ordering among categories is not considered. At present, only a few ordinal-sensitive procedures for learning Bayesian network from ordinal data have been developed in order to preserve the ordering of ordinal data (Musella, 2013).

Following Musella's procedure (Musella, 2013), we describe an ordinal PC algorithm for learning and inferring a QTL DAG from marginal ranked genotypic values. Consider data structure, given in **Table S14** (right), composed of eight samples on six QTL variables. The PC algorithm is a stepwise backward algorithm for DAG inference (Spirtes et al., 2013). We

first tested the conditional independence  $Q_1 \perp Q_3 | Q_2$  where  $Q_1$ ,  $Q_2$  and  $Q_3$  are ordinal, each with three genotypes. We let  $n_{j_1 j_2 j_3}$  denote the observation of the  $j_1$ -th genotype of  $Q_1$  ( $j_1 = 1, 2, 3$ ),  $j_2$ -th genotype of  $Q_2$  ( $j_2 = 1, 2, 3$ ), and  $j_3$ -th genotype of  $Q_3$  ( $j_3 = 1, 2, 3$ ). Let  $F_{j_1 j_2}(Q_3)$  denote the conditional distribution of  $Q_3$  given  $Q_1 = j_1$  and  $Q_2 = j_2$ . The null hypothesis of the test is formulated as

$$H_0: F_{1j_2}(j_3) = F_{2j_2}(j_3) = F_{3j_2}(j_3), \forall j_2, \forall j_3 \quad (8)$$

whose alternative test reflecting a stochastic ordering among distributions is written as

$$H_1: \begin{cases} F_{j_1 j_2}(j_3) > F_{j'_1 j_2}(j_3) \\ F_{j_1 j_2}(j_3) < F_{j'_1 j_2}(j_3) \end{cases}, \text{ with } j_1 < j'_1, \forall j_2, \forall j_3 \quad (9)$$

From tests (8) and (9), we calculate a so-called Jonkheere-Terpstra (JT) test statistic as

$$JT = \sum_{j_2=1}^3 \sum_{j_1=1}^3 \sum_{j'_1=1}^{j_1-1} \left[ \sum_{s=1}^3 w_{j_1 j'_1 s j_2} n_{j_1 s j_2} - \frac{n_{j_1 \cdot j_2} (n_{j_1 \cdot j_2} + 1)}{2} \right] \quad (10)$$

where  $w_{j_1 j'_1 s j_2}$  was the Wilcoxon score described by

$$w_{j_1 j'_1 s j_2} = \sum_{t=1}^{s-1} (n_{j_1 t j_2} + n_{j'_1 t j_2}) + \frac{n_{j_1 s j_2} + n_{j'_1 s j_2} + 1}{2}. \quad (11)$$

Under the null hypothesis, the mean of JT was calculated as

$$E(JT|H_0) = \frac{1}{4} \sum_{j_2=1}^3 (n_{\cdot \cdot j_2}^2 - \sum_{j_1=1}^3 n_{j_1 \cdot j_2}^2)$$

Per Lehmann (D'Abrera and Lehmann, 1975) and Pirie (Pirie, 1983), we derive the asymptotic variance of JT under the null hypothesis as

$$\widehat{Var}(JT|H_0) = \frac{V_1}{72} + \frac{V_2}{36(n_{\cdot \cdot j_2} - 1)(n_{\cdot \cdot j_2} - 2)} + \frac{V_3}{8(n_{\cdot \cdot j_2}(n_{\cdot \cdot j_2} - 1))}$$

where

$$\begin{aligned} V_1 &= n_{\cdot \cdot j_2} (n_{\cdot \cdot j_2} - 1) (2n_{\cdot \cdot j_2} + 5) - \sum_{j_3}^3 (n_{j_1 \cdot j_2} (n_{j_1 \cdot j_2} + 5) - \sum_{j_3}^3 (n_{j_2 j_3} - 1) (2n_{j_2 j_3} + 5)), \\ V_2 &= \sum_{j_1}^3 (n_{j_1 \cdot j_2} (n_{j_1 \cdot j_2} - 1) (2n_{j_1 \cdot j_2} - 2) - \sum_{j_3}^3 (n_{j_2 j_3} - 1) (n_{j_2 j_3} - 2)), \\ V_3 &= \sum_{j_1}^3 (n_{j_1 \cdot j_2} (n_{j_1 \cdot j_2} - 1)) - \sum_{j_3}^3 (n_{j_2 j_3} (n_{j_2 j_3} - 1)). \end{aligned}$$

Based on these derivations, it can be proved that the test statistic is asymptotically normally distributed.

The ordinal PC algorithm infers a QTL DAG (Musella, 2013) using the three steps as follows:

**Step 1: Build the skeleton of the graph.** Starting with a complete undirected graph, we obtain a graph where all QTL are connected to each other. Given a chosen significance level, statistical tests based on the Jonkheere-Terpstra test are performed to decide if we need to remove or maintain edges between QTL in the graph. This procedure leads to the detection of the skeleton of the graph.

**Step 2: Find v-configurations.** If two QTL,  $Q_1$  and  $Q_2$ , are not conditionally independent given a QTL  $Q_3$ , then  $Q_3$  is a collider node or a common sink and a v-configuration  $Q_1 \rightarrow Q_3 \leftarrow Q_2$  (i.e., converging directed edges into the same node) is drawn; otherwise edges remain undirected  $Q_1 - Q_3 - Q_2$ .

**Step 3: Create no new v-configuration.** Some constraints must be given to orient other edges without creating additional colliders or some cycles.

Musella (2013) showed that the ordinal PC algorithm outperforms the PC algorithm (for modeling discrete data without considering their ranking) in terms of sensitivity (specified by true positive rate) and precision (specified by true discovery rate), especially when sample size is small. Yet, the two algorithms do not differ dramatically in specificity (specified by false positive rate).

**Social network:** Different animals interact with each other through mutualism, antagonism, aggression, or altruism to form a community. We implement an ordinary differential equation (ODE) approach proposed by Wu et al. (2014) to investigate how QTL modulate the structure and organization of an animal-animal interaction network. In such a QTL-driven social network, the nodes are individual animals and the edges are animal-animal interactions whose direction, sign, and strength are determined by QTL. Consider **Table S14**'s toy example for a mapping population, where eight animals were each collected by six markers. We assign each individual at a given QTL by a value, i.e., its marginal genotypic mean at this QTL. Taken together, we obtain an  $(8 \times 6)$  matrix of genotypic values. Unlike a QTL each with three ranked categories of genotypic values over all animals, each animal form a set of somewhat continuous genotypic values across QTL (right, **Table S14**). Next, we show that a nonparametric approach can be used to model how an animal changes its genotypic value over QTL.

Let  $g_{ik}$  denote the genotypic value of individual  $i$  ( $i = 1, \dots, 8$ ) at QTL  $k$  ( $k = 1, \dots, 6$ ). Note that  $g_{ik}$  depends on the genotype individual  $i$  carries at QTL  $k$ . Because of social interactions, the genotypic value of one individual is affected by or affects those of other individuals. This allows us to formulate a system of ODEs, expressed as

$$\dot{g}_{ik} = f_i(g_{ik}) + \sum_{i' \neq i, i'=1}^8 h_{i \leftarrow i'}(g_{i'k}) \quad (12)$$

where  $\dot{g}_{ik}$  is the rate of the overall change of genotypic value for individual  $i$  from one QTL to next,  $f_i(g_{ik})$  is the function that describes the change rate of individual  $i$ 's genotypic value independent of any other individuals, and  $h_{i \leftarrow i'}(g_{i'k})$  is the function that specifies the change rate of genotypic value due to the impact of any other individual  $i'$  that affects individual  $i$ .  $h_{i \leftarrow i'}(g_{i'k})$  determines the sign, direction, and strength of social interaction between individual  $i$  and  $i'$ . If both  $h_{i \leftarrow i'}(g_{i'k})$  and  $h_{i' \leftarrow i}(g_{ik})$  are positive or negative, this suggests that these two individuals are mutualistic and antagonistic, respectively. If both are zero, then the two individuals have no interactions. If  $h_{i \leftarrow i'}(g_{i'k})$  is positive or zero but  $h_{i' \leftarrow i}(g_{ik})$  is negative, this indicates that individual  $i$  is aggressive on individual  $i'$ . If  $h_{i \leftarrow i'}(g_{i'k})$  is positive but  $h_{i' \leftarrow i}(g_{ik})$  is zero, this shows that individual  $i'$  is altruistic for individual  $i$ . The magnitudes of  $h_{i \leftarrow i'}(g_{i'k})$  and  $h_{i' \leftarrow i}(g_{ik})$  can quantify the strength of social interactions.

To solve ODE (12), we integrate three rules from different disciplines. First, in sociological studies, there is the Dunbar's law, stating that the number of stable relationships a human can comfortably maintain in his social network is not beyond a limit (Dunbar, 1992). Under the Dunbar's law, using real data from both hunter-gather and modern communities, Harre and Prokopenko (2016) calculated the average number of links maintained by individuals to form cooperative groups, which changes with group size. For example, a person in a group of five,

15, 45, and 132 needs to maintain an average of one to two, two to three, three to four links, and four to five links respectively. This rule can largely simplify our joint modeling of ODEs (12), making ODE parameter estimation and interpretation feasible and tractable, especially when the dimension of this equation group is large. Second, statistical rules of variable selection can be used to determine an optimal small set of individuals that interact stably with a focal individual. Group LASSO (Yuan and Lin, 2006) and adaptive group LASSO (Wang and Leng, 2008) derived from Tibshirani's original LASSO (Tibshirani, 1996) have proven to be powerful for variable selection. In previous studies, several authors have already worked out model selection in high-dimensional ODEs (Lu et al., 2011; Henderson and Michailidis, 2014; Wu et al., 2014). Thus, the integration of the Dunbar's law with variable selection enables the construction of a sparse social network, which facilitates the characterization of most important social connections within the network.

Third, how does the locus-varying genotypic value of an individual change from one QTL to next? In quantitative genetics, the genetic effect of a QTL is generally positively associated with its genetic variance. Based on this rule, we can model the genotypic value of an individual at a QTL as a function of standard genetic deviation explained by this QTL. By arranging all QTL in an order of their genetic deviations, we incorporate a Legendre Orthogonal Polynomial (LOP)-based nonparametric approach to fit the functions,  $f_i(g_{ik})$  and  $h_{i \leftarrow i'}(g_{i'k})$ , that jointly describe QTL-varying genotypic value of each individual in equation (12). Because of its advantage in orthogonality and efficient convergence, the LOP is effective for modeling the curves of any complex form using sparse data in quantitative genetic studies (Das et al., 2011; Jiang et al., 2016). The LOP, a solution of the Legendre differential equation,

$$(1 - v^2) \frac{d^2 u}{dv^2} - 2v \frac{du}{dv} + r(r + 1)u = 0$$

can be expressed as

$$P_r(v) = \sum_{c=0}^r (-1)^c \frac{(2r - 2c)!}{2^r! (r - c)! (r - 2c)!} v^{r-2c}$$

which is called the Legendre polynomial of order  $r$ , where  $C$  is an integer, expressed as  $r/2$  or  $(r - 1)/2$ , and  $v$  is an independent variable, i.e., standard genetic deviation in social network modeling. In practice, it needs to be corrected as  $t^* = 1 + 2(t - t_{\max})/(t_{\max} - t_{\min})$  within interval  $[-1, 1]$ , where  $t_{\min}$  and  $t_{\max}$  are the two extreme points at the low and high end, respectively. By defining a series of basis values, the LOP is used to determine the curvature of QTL-varying genotypic values by choosing an optimal polynomial order.

In **Table S14** (right), we construct the QTL network and social network for the toy example by the above approaches. These toy networks can help the readers better understand the utility of these approaches.

### Monte Carol simulation

To examine the statistical properties of the new model, we perform computer simulation by mimicking the data structure of a mapping population. We show how to simulate the phenotypic data of a trait under the constraint of animal-animal interactions. We let  $w_{j_1}$  and  $w_{j_2}$  denote the phenotypic value of animal  $i_1$  and  $i_2$  ( $0 \leq i_1 < i_2 \leq m$ ), respectively, from the population of  $m$  animals and  $z_i$  ( $i = 1, \dots, m(m-1)/2$ ) denote the strength of one of their interaction types, as defined in Fig. 1. To simulate the trait data of the pairing animals, we need to determine the variance of phenotypic values among  $m$  animals under the constraint of a social interaction considered. This requires us to derive the variance of an interaction derivative

( $z_i$ ) among  $m(m-1)/2$  pairs in terms of the variances of  $w_{j_1}$  and  $w_{j_2}$ . As an example, we assume that the aggression variable, expressed as the ratio of a larger to smaller animal, is simulated.

We reformat the data by adding reciprocal pairs, generating a total of  $m(m-1)$  pairs. Let  $x > y$  denote the phenotypic values of animals at the left and right sides, respectively, in pairs. Assuming that  $x$  and  $y$  are independent, the variance of  $z = x/y$  is expressed by

$$V_Z = \frac{1}{\bar{y}^4} (\bar{y}^2 V_X + \bar{x}^2 V_Y - 2\rho\bar{x}\bar{y}\sqrt{V_X V_Y}), \quad (13)$$

where  $\bar{x}$  and  $\bar{y}$  are the means of  $x$  and  $y$  variables,  $V_X$  and  $V_Y$  are the variances of  $x$  and  $y$  variables, and  $\rho$  is the correlation between the two variables, respectively. In spite of different orders,  $x$  and  $y$  variables contained the same set of phenotypic data for  $m$  fish so that  $V_X = V_Y$  and  $\bar{x} = \bar{y}$ .

Consider a testcross QTL with two genotypes  $AA$  and  $Aa$  of an equal proportion in the mapping population. A total of  $m(m-1)$  pairs are clustered into four GG combinations (right panel, **Table S12**). The phenotypic data of an animal pair are simulated by summing the GG genotypic value and a residual error with mean zero and residual variance ( $V_Z$ ) (14) scaled by the heritability ( $H_Z^2$ ) explained by GG combinations at the assumed QTL. From the simulated data of animal pairs, we need to simulate the phenotypic data of individual animals (left, **Table S12**). This was done by randomly sampling values of  $m$  animals that meet their ratios across  $m(m-1)/2$  pairs, but under the constraint that  $m$  animals' values have a variance  $\sqrt{V_Z}$ , expressed by equation (13).

Our simulation was based on four scenarios designed per GG heritability  $H_Z^2$  (large 0.10 vs. small 0.05) and mapping size  $m$  (large 200 vs. small 70). For each scenario, we calculated the accuracy and precision of each genetic effect (such as the direct, indirect, and genome-genome epistatic effects) from pairwise data (right, **Table S12**) by the new model, testing the statistical property of the new model. Also, we calculate and compare the power of QTL detection by the new model and the traditional model that analyzes original data directly (left, **Table S12**). The false positive rates of the new model is also calculated.

### Mapping experiment

**Animal material:** The experiment of genetic mapping was conducted using an  $F_1$  family (named H1) of *Cyprinus carpio* including 71 progeny produced by Hebao Red carp and Koi carp. The fish were cultured at the Research Institute for Heilongjiang River Fisheries, Harbin, China, and measured for body mass after anesthesia with MS222 when they reached an adult stage of fish growth. The H1 family was served as the discovery cohort, whose results were directly used to test our new theory. To validate the results discovered by the new theory, we replicated the mapping study by two additional  $F_1$  families (named G1 and Z22) of Yellow River carp. These two families with 115 and 62 progeny, respectively, were cultured at the Henan Academy of Fishery Sciences, Zhengzhou, China. The same trait, body mass, was measured for each family at the adult stage of fish growth.

**SNP array genotyping and quality control:** Genomic DNA was extracted from blood samples (400–800 $\mu$ l caudal peduncle) of the hybrids and their parents using a DNeasy Blood & Tissue Kit (Qiagen, Shanghai, China) following manufacturer's protocol. DNA was quantified by Nanovue (Thermo Scientific) and the integrity of DNA was examined by 1% agarose gel electrophoresis stained with ethidium bromide. Then qualified DNA was

genotyped using the Affymetrix Axiom Carp SNP array containing ~250 K validated SNPs (P. Xu et al., 2014; J. Xu et al., 2014). Genotyping results were provided by GeneSeek (Lincoln, Nebraska, USA). After quality control, we obtained 39,960 Mendelian segregating SNPs throughout the common carp genome of size ~1.42 Gb in the H1 family, from which a high-density linkage map, with an average marker interval of 0.75 cM, was constructed by OneMap (Margarido et al., 2007). For G1 and Z22 families, we genotyped 39,960, 97,532, and 86,370 SNPs following Mendel's first law, respectively.

**Annotation of candidate genes:** The significant SNPs detected by our theory are annotated using the database of the common carp genome project. Genes located within upstream and downstream 10 kb of the candidate SNPs were selected. Based on the selected gene lists, Gene Ontology and KEGG Enrichment analyses are conducted using DAVID (Huang et al., 2009; Huang et al., 2009) online programs with default parameters, and figures are drawn using REVIGO software (Supek et al., 2011).

## References

- Cedilnik, A., Kosmelj, K. and Blejec, A. (2004). The distribution of the ratio of jointly normal variables. *Metodol Zvezki* 1, 99-108.
- Das, K., Li, J.H., Wang, Z., Fu, G., Li, Y., Mauger, D., Li, R. and Wu, R.L. (2011). A dynamic model for genome-wide association studies. *Human Genetics* 129, 629–639.
- Dunbar, R.I.M. (1993). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22, 469–493.
- Harre, M. and Prokopenko, M. (2016). The social brain: Scale-invariant layering of Erdos-Renyi networks in small-scale human societies. *Journal of the Royal Society Interface* 13, 1-6.
- Henderson, J. and Michailidis, G. (2014). Network reconstruction using nonparametric additive ODE models. *PLoS ONE* 9(4), e94003.
- Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* 37, 1–13.
- Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4, 44–57.
- Jiang, L.B., Ye, M.X., Zhu, X.L., Sang, M.M. and Wu, R.L. (2016). Evo-Devo-EpiR: A genome-wide search platform for epistatic control on the evolution of development. *Briefings in Bioinformatics* 18, 754-760.
- D'Abrera, H.J.M. and Lehmann, E.I. (1975). Nonparametrics: Statistical Methods Based on Ranks. Holden-Day.
- Lu, T., Liang, H., Li, H. and Wu, H. (2011). High-dimensional ODEs coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *Journal of American Statistical Association* 106, 1242–1258.
- Lu, Q., Cui, Y.H. and Wu, R.L. (2004). A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. *BMC Genetics* 5, 20.
- Margarido, G.R.A., Souza, A.P. and Garcia, A.A.F. (2007). OneMap: Software for genetic mapping in outcrossing species. *Hereditas* 144, 78–79.
- Musella, F. (2013). A PC algorithm variation for ordinal variables. *Computational Statistics* 28, 2749–2759.
- Pirie, W. (1983). Jonckheere tests for ordered alternatives. *Encyclopedia of Statistical Science* 4, 315–318.
- Spirtes, P., Glymour, C. and Scheines, R. (2000). Causation, Prediction, and Search, 2nd edn, MIT Press.

- Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7), e218000.
- Tibshirani, R.J. (1996). Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society B* 58, 267-288.
- Wang, H. and Leng, C. (2008). A note on adaptive group LASSO. *Computational Statistics and Data Analysis* 52, 5277-5286.
- Wu, H., Lu, T., Xue, H. and Liang, H. (2014). Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. *Journal of American Statistical Association* 109, 700-716.
- Wu, R., Ma, C., Painter, I. and Zeng, Z. (2002). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theoretical Population Biology* 61, 349-363.
- Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., Kuang, Y., Zhang, Y. et al. (2014). Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics* 15: 307.
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y. and Zhang, Y. (2014). Genome sequence and genetic diversity of the common carp. *Nature Genetics* 46, 1212-1219.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of Royal Statistical Society B* 68, 49-67.
- Zhu, X., Jiang, L., Ye, M., Sun, L., Gragnoli, C. and Wu, R.L. (2016). Integrating evolutionary game theory into mechanistic genotype-phenotype mapping. *Trends in Genetics* 32, 256-268.