Spring 2019 – Epigenetics and Systems Biology
Discussion Session (Epigenetics and Evolutionary Biology)
Michael K. Skinner – Biol 476/576
Week 16 (April 25)

Epigenetics and Evolutionary Biology

Primary Papers

1. Soyer and O'Malley (2013) Bioessays 35:696-705
2. Skinner, et al. (2014) BMC Genomics 15:377
3. Artemov, et al. (2017) Mol. Bio. Evol. 34(9):2203-13
4. McNew, et al. (2017) BMC Evolution 17:183

Discussion

Student 45 – Ref #1 above

- What is the evolutionary systems biology referred to?
- How can systems biology help understand evolution?

Student 46 – Ref #2 above

- What is sexual selection?
- How did epigenetic transgenerational inheritance change mate preference?
- Does this provide evidence for environmental induction of epigenetic transgenerational inheritance and a role in evolutionary biology?

Student 47 – Ref #3 above

- What is the model system and phenotypic change?
- What epigenetic change was observed?
- How did the epigenetics and genetics integrate?

Student 48 – Ref #4 above

- What was the experimental design and approach?
- What molecular alterations were observed in what cell types?
- What molecular mechanism can promote rapid evolutionary events?

Evolutionary systems biology: What it is and why it matters

Orkun S. Soyer[1] and Maureen A. O'Malley[2]

Evolutionary systems biology (ESB) is a rapidly growing integrative approach that has the core aim of generating mechanistic and evolutionary understanding of genotype-phenotype relationships at multiple levels. ESB's more specific objectives include extending knowledge gained from model organisms to non-model organisms, predicting the effects of mutations, and defining the core network structures and dynamics that have evolved to cause particular intracellular and intercellular responses. By combining mathematical, molecular, and cellular approaches to evolution, ESB adds new insights and methods to the modern evolutionary synthesis, and offers ways in which to enhance its explanatory and predictive capacities. This combination of prediction and explanation marks ESB out as a research manifesto that goes further than its two contributing fields. Here, we summarize ESB via an analysis of characteristic research examples and exploratory questions, while also making a case for why these integrative efforts are worth pursuing.

Keywords:
evolution; genotype-phenotype mapping; integration; molecular networks; system dynamics

DOI 10.1002/bies.201300029

[1] Warwick Centre for Synthetic Biology, School of Life Sciences, University of Warwick, Coventry, UK
[2] Department of Philosophy, University of Sydney, Sydney, NSW, Australia

Corresponding authors:
Orkun S. Soyer
E-mail: o.s.soyer@exeter.ac.uk
Maureen A. O'Malley
E-mail: maureen.omalley@sydney.edu.au

The emergence of ESB

At its most basic, evolutionary systems biology (ESB) is the synthesis of system-level approaches to biological function with evolutionary explanations of multilevel properties. "System" in this context refers to dynamically interacting components that produce behavior not revealed by analyses of isolated components. Cellular interactions of signaling, regulatory, and metabolic components are all considered as systems or networks that can display structural intricacies and nonlinear dynamics. ESB recognizes that the system-level properties of cellular networks are subject to evolutionary change, and that evolved network properties will variously influence the future evolutionary course of the organism. It is this interdependency between evolutionary processes and system properties that ESB aims to understand.

One of the earliest articles to describe ESB was published in 2005 [1]. While this overview and subsequent elaborations (e.g. [2]) were milestones in the development of ESB as a research programme, their scope was restricted to comparative analyses. Since 2005 ESB has flourished, to the extent that comparative methods currently constitute only one facet of its diverse research strategies (in fact, many discussions of ESB prefer to exclude this comparative work in favor of more dynamic accounts; we make a case for including it below). ESB also goes beyond existing efforts to merge molecular and evolutionary biology (e.g. [3]). Those projects aim to understand evolutionary processes in the narrower context of individual genes and structure of their protein products, whereas ESB is committed to studying phenotypes as the results of evolving intracellular interaction networks (Fig. 1). At present, only a few ESB researchers refer self-consciously to their work with the label "evolutionary systems biology" (see [4]). However, this situation is changing as increasing numbers of evolutionarily minded, systems-focused researchers forge connections between a range of research questions (e.g. [5, 6]).

In contrast to some views of ESB, which suggest it is defined by a specific research agenda [1, 2, 7], we see the distinguishing capacity of ESB as its ability to integrate theoretical tools, experimental methods, and extensive datasets within an evolutionary framework. This integration is

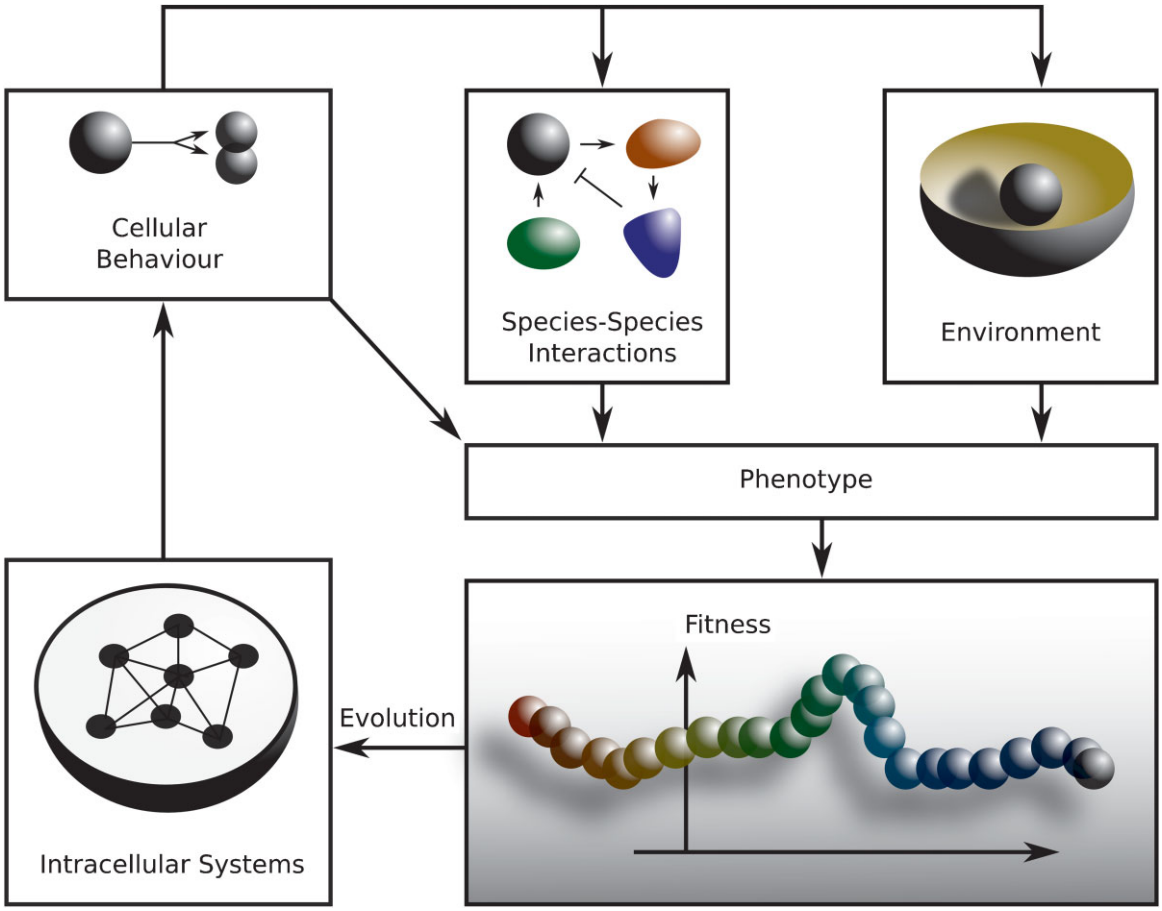

Figure 1. Systems under evolutionary forces. At the core of ESB lies the aim of achieving a deep mechanistic understanding of genotype-phenotype mappings in biological systems. While these mappings can be drawn at different levels in different combinations of ESB, a major area of interest is currently intracellular systems. These systems give rise to cellular physiology, which – in the case of unicellular organisms – directly determines species' interactions with their environment and other organisms. These higher-level interactions are responsible for the fitness of organisms. Evolutionary processes (i.e. neutral drift and adaptation) move populations of these organisms on this dynamic fitness landscape by altering the properties of their intracellular systems. Credit: Arno Steinacher created this figure.

occurring in a highly pragmatic manner, and is driven by the core ambition of ESB practitioners to develop closer insight into evolving genotype-phenotype mappings across different biological scales (Fig. 1). Researchers with this goal seize upon tools and datasets as they become available (e.g. dynamical models, gene-knockout studies, flux balance analyses, in silico evolution, reverse engineering, comparative omics data; Fig. 2) to address questions as old as biology or to reformulate new ones in light of system-level insight.

Some readers may find this description of ESB too methodologically inclusive. It could be argued that to be a viable research approach, ESB should be more restricted in what it does and does not do. But methodological flexibility and a wide investigative agenda are what characterize many new movements in biology (e.g. systems biology itself); historically, molecular biology was not the product of a precisely organized research agenda [8]. It seems unlikely to us that biological research programmes come about as the planned products of rational intellectual design and social engineering. Instead, they emerge from communities of researchers as piecemeal responses to the needs and constraints experienced in existing fields of research. ESB is such a product. It may indeed resist any tight definitions, but as we will show, it has a synthesizing mission that distinguishes it from its contributing fields.

While systems biology may have initially overlooked evolution because modeling network dynamics is already demanding enough, there are also explicit arguments *against* taking an evolutionary approach to systems biology. One of these is that to understand the organization of systems, an evolutionary perspective is irrelevant. While specific organizational features may have evolved to cope with environmental perturbations, it is getting a grip on "how" these are implemented at the molecular level rather than "why" that is considered the important task (e.g. [9]). We will address this objection below by showing that organizational properties have dimensions appropriately and necessarily addressed by ESB. Another argument for leaving evolutionary inquiries out of systems biology might be that population genetic approaches are sufficient to understand mutations and their effects. Again, we will show how the combination of

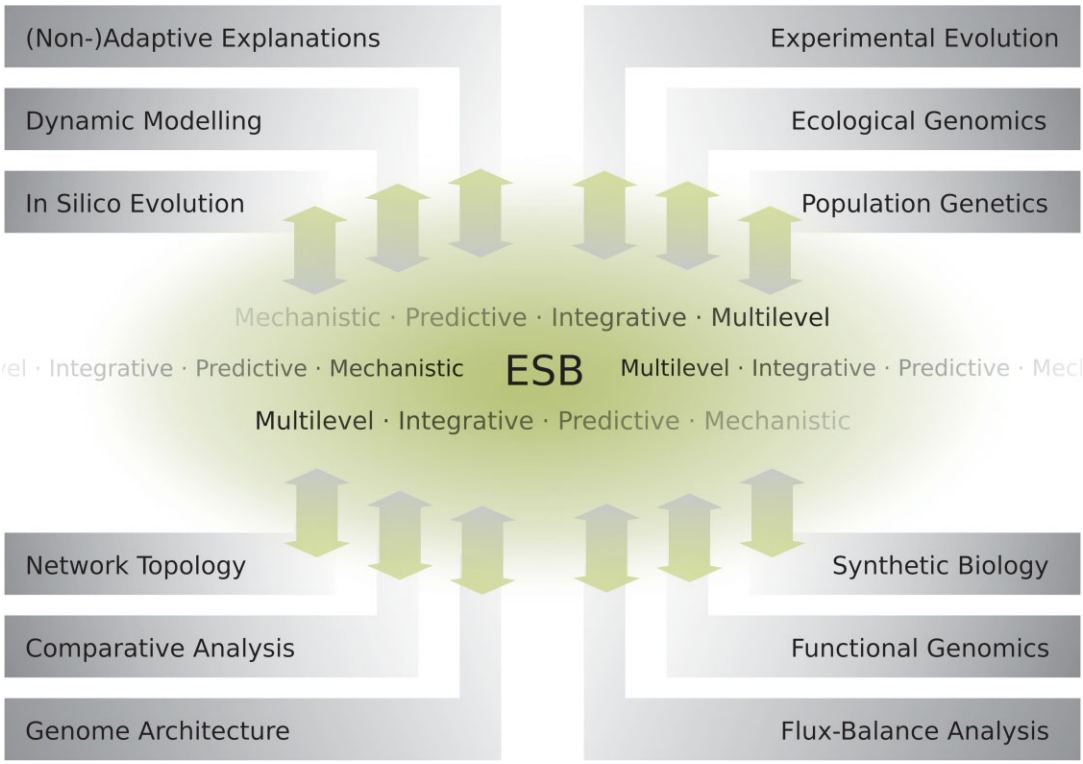

Figure 2. A representation of methodological inputs into ESB. There is great variety in the methods and perspectives that can and do contribute to ESB, and this figure is merely indicative rather than an exhaustive list of contributions. The examples outlined in the text depict various combinations of these methodological inputs, although no single ESB study is likely to combine all of them. It is important to note that the interaction between ESB and these methodological contributions is two-way, with new insights and innovations achieved by ESB having feedback effects on the contributing tools and methods. ESB itself is conceptualized as an open interdisciplinary space produced by the integration of these various approaches with evolutionary frameworks. Four ESB potentials are highlighted in that space: mechanistic, predictive, integrative and multilevel capacities. Credit: Arno Steinacher created this figure.

evolutionary and systems biological analyses reveals further dimensions of mutational processes and leads to better understanding of their evolutionary outcomes.

ESB research questions

Despite the diversity of methods used in ESB, there is a broadly unifying motivation that drives the approach. The most general reason for why systems biology requires an evolutionary framework, and why evolutionary biology should concern itself with systems analyses, is as we have suggested above: each will be a more complete science with the other perspective taken into account. ESB integrates "how" and "why" questions about biological systems, and thus appeals to network biologists who want to understand not just how a system functions but why it functions in a particular way. This combination of predictive and explanatory capacities is a distinct feature of ESB and we will elaborate on it below. But more specifically, ESB has a number of focused research objectives that set the agenda for different aspects of its practice, and it is these objectives that determine methodological combinations. We will outline a range of these questions in relation to how they can be answered.

Generalizations of network properties beyond model species

Usually, any systems-biological understanding of network dynamics and structure is gained via modeling and experimentation with regard to a particular model system. For example, network modeling and experiments have shown that signaling networks enabling chemotaxis in *Escherichia coli* and osmoregulation in *Saccharomyces cerevisiae* display transient (i.e. adaptive) dynamics [10, 11]. While these and similar studies show how a particular physiological response is enabled by specific system dynamics and structure, they cannot easily be generalized to other organisms. A whole raft of questions remains unanswered: Are the observed system dynamics necessary for the physiological response? What are alternative molecular structures that have the same system dynamics? Could these alternatives be realized under different ecological and evolutionary pressures? Answering these questions requires going beyond organism-specific systems biology with the help of evolutionary analyses. The evolutionary methods that are used to answer such questions include experimental and mathematical comparison of network architecture and dynamics in different species (e.g. [12–14]), comparative genomic analyses at the network

level (e.g. [15, 16]), in silico evolution of network models (e.g. [17, 18]), and statistical inference and network reconstruction methods (e.g. [19]).

Determining the evolutionary forces behind network-level patterns

Numerous studies have demonstrated global or recurring features in cellular networks (e.g. [20]). These organizational characteristics exist at different levels of the biological hierarchy and include specific connectivity distributions, the presence of modules and motifs, and biochemical features (e.g. cooperative binding, noise, and multifunctional enzymes). Understanding the significance of these features requires an account of the evolutionary forces that shaped the networks in which these patterns were found. In particular, the observed features can be adaptive (e.g. the result of specific selective pressures such as stabilizing or fluctuating selection [21]) or neutral (e.g. resulting from genetic drift [22]), or could have emerged due to a mix of these two forces (i.e. neutral mutations that become adaptive in new circumstances) [23]. Methods include in silico evolution of simple network models under adaptive and neutral forces, comparison of network features from different species (e.g. with different population sizes giving different roles to genetic drift, or different selection pressures producing different network features), and population genetics.

Comprehensive understanding of mutational effects

Another theme in systems biology is predicting the effects of mutations on particular functions of cellular networks (e.g. large-scale mutation or deletion studies such as those undertaken in *S. cerevisiae* [24]). Evolutionary biology, on the other hand, aims to discern the role of mutational effects in any evolutionary process. In general, systems biology addresses this research theme within the model-experiment cycle, while evolutionary biology addresses it by creating simplified models and evaluating their ability to capture experimental observations [25]. ESB bridges these two methodologies to produce a more comprehensive understanding of mutational effects. For example, sensitivity analysis over different network models has been used to gain insight into the distribution of mutational effects on a specific property of system dynamics, such as the phase of oscillation in a circadian clock model [7, 26]. In the case of metabolism, mathematical models of optimal metabolite flows in metabolic networks under a steady state assumption (metabolic flux-balance analysis (FBA [27]), have been used to predict phenotypes produced by single-enzyme deletions [6]. Experimental evolution can be combined with next-generation sequencing to characterize both the mutations and their fitness effects (e.g. antibiotic resistance [28]). A valuable methodological development in ESB would be establishing a cycle between evolutionary experimentation and system-level modeling.

Extending systems biology to co-evolved intercellular interactions

Contemporary molecular systems biology has until recently focused on intracellular networks. Examining intercellular network interactions in homogenous cultures of microorganisms (and between cells of multicellular organisms) has been carried out in the context of metabolism (via exchanged metabolites), quorum sensing, and the production of "public goods" such as scavenging enzymes or co-factors. In host-parasite interactions, heterogeneous cellular interactions can occur via parasite-encoded genes that function in the networks of the host. Systems-biological study of these cellular interactions does not necessarily require an evolutionary framework (e.g. applications of FBA to interactions between two species [29, 30]). However, cellular interactions have historically been studied to great effect in an evolutionary context, particularly with regard to the co-evolutionary dynamics and evolution of social behavior in microbes [31]. ESB is strategically well placed to combine network approaches with these broader evolutionary perspectives.

To sum up, the research questions driving ESB combine those asked separately in evolutionary biology and systems biology. In addressing these questions, ESB integrates the tools and approaches of evolutionary biology and systems biology to achieve results that are more than additive: systems biology gains a more multilevel, explanatory perspective with the incorporation of evolutionary timescales, while evolutionary biology is made more quantitative, predictive, and able to combine mathematical abstractions with mechanistic detail. ESB research can therefore deliver insights not possible from evolutionary or systems biology approaches applied in isolation. We will show more specifically how this works with two apt illustrations of ESB's constructive dynamic.

ESB exemplified: Robustness, epistasis, and pleiotropy

While there are large numbers of ESB studies to choose from (e.g. [5]), we will use as exemplars two areas of research that not only illuminate ESB-in-action, but also deal with some of the resistance to ESB we mentioned above. Robustness refers to a system's ability to maintain its output in the face of perturbations. It has been demonstrated for several phenotypes via gene knockout studies [24] and metabolic flux measurements [32]. These elegant systems-biological studies were descriptive, however, and could not explain the evolutionary significance of robustness; nor could they make predictions about which evolved systems should be expected to display this property [33]. While several theoretical studies have attempted to address these questions (e.g. [34, 35]), more complete insights have come from recent ESB work that amalgamates population genetics, network-level dynamical models, genomic analyses, flux-balance models, and in silico evolution.

These studies point to plausible selective pressures that would generate robustness against gene deletion in gene-regulatory and metabolic networks [36–39]. Selection for

stable gene-expression patterns under recombination results in the evolution of robustness as a product of network connectivity [40] and the emergence of epistatic interactions among nodes [36, 37]. In the case of metabolic networks, selection for biomass production from variable resources (i.e. fluctuating selection) leads to the evolution of robust networks with overlapping metabolic routes and multifunctional enzymes [38]. These mechanistic determinants of robustness fit closely the findings of flux-redistribution experiments [32]. Additional ESB studies combining genomic data and FBA add support to claims that robustness depends on environmental conditions [39, 41].

These novel insights into the mechanistic determinants and evolutionary drivers of network-level robustness came about by synthesizing the methods of systems and evolutionary biology. Mechanistic findings complement systems-biology studies that aim to decipher the overall structure, dynamics, and function of metabolic networks (e.g. [32]). The improved understanding of the evolutionary forces involved has generated valuable predictions, including the hypotheses that gene-regulatory networks harbor genes able to buffer the effects of mutations [36], and that metabolic robustness in microbes parallels the evolutionary and ecological exposure of those organisms to fluctuating environments [38, 39, 41].

Epistasis (multiple genes or mutations producing phenotypic effects that deviate from a simple additive model) and pleiotropy (in which one gene or mutation affects multiple phenotypic traits) are crucial aspects of genotype-phenotype relationships [42, 43]. In ESB, the systematic application of FBA to genome-scale metabolic models of yeast reveals that epistasis commonly takes place between functional modules of genes [44]. This finding shifts the focus from interactions occurring between individual genes to those occurring between modules. Subsequent expansion of the same analysis to many metabolic environments and multiple phenotypes shows that gene pairs in yeast can interact "incoherently" relative to different phenotypes (i.e. negatively in relation to some phenotypes and positively in relation to others) [45]. These discoveries would have been invisible to any perspective based solely on individual interactions or simple models. Similarly, studies of system-level models of developmental regulatory networks indicate that both the overall function of a gene-regulatory network [46] and its gene-regulatory patterns [47] can give rise to specific signatures of statistical epistasis. The latter finding is potentially of high methodological importance to evolutionary biology because it shows that combining models of nonlinear network dynamics with commonly used genetic methods, such as quantitative trait loci mapping, can greatly improve the performance of each approach [47, 48].

Network-level research into epistasis often inevitably leads to insight into pleiotropic effects and interactions [49, 50]. Investigations of genome-wide pleiotropic effects in yeast, nematodes, and mice indicate that pleiotropy, while occurring at a low rate, is highly modular and has an important evolutionary role in the generation of adaptive complexity in organisms [51] (however, see [52]). Studies focused on single adaptive mutations with whole-network effects thus offer general insight into pleiotropy as a mechanism of evolution [53]. Ongoing ESB research seeks to clarify how epistasis and pleiotropy enable or obstruct evolutionary change at the network level, and whether these processes are simply inevitable byproducts of network connectivity [45, 43].

From old to new questions in ESB

ESB is able to elaborate on existing answers to old evolutionary questions as well as devise new system-level evolutionary questions. An example of the former, in which existing questions are addressed by new ESB approaches, is found in the fields of comparative genomics and evo-devo. An ESB extension of classic comparative genomics focuses on complete systems of interacting genes to decipher the variance in the structure of these networks (i.e. what interacts with what) and the evolutionary history of those interactions [54, 16]. These comparisons also take into account the evolutionary dynamics of both genomic and cellular network architecture [55–57].

The field of evo-devo has profited considerably from contemporary capacities for comparative network analysis, especially in regard to understanding the relationships between developmental processes and the evolution of transcriptional regulatory networks (notably *cis*-regulatory networks) [58–60]. The integration of new datasets now enables the incorporation of additional levels of regulation in the study of evolving developmental processes. In particular, it is becoming increasingly feasible to examine the roles of post-transcriptional and post-translational regulation in development [61]. Emerging ESB studies are already trying to study these processes and their connections [62, 63].

But perhaps a more thoroughgoing integration is happening in regard to the combination of comparative approaches with theoretical tools such as dynamical modeling, mathematical inference, and in silico evolution [64]. This amalgamation has already allowed the prediction of underlying regulatory networks from the analysis of phenotypic (i.e. gene expression) data in the model organism *Drosophila* [12]. Some evo-devo proponents have argued that modeling these networks dynamically merely confirms the results of existing qualitative experimental approaches but adds no additional insight (e.g. [65]). However, models that incorporate system dynamics, network evolution, and phenotypic datasets have the advantage of being able to examine directly genotype-phenotype mapping in diverse organisms [14, 66, 67], which is something qualitative experimentation cannot do. These models have revealed that due to nonlinear interactions, knowledge of network structure is often *not* sufficient for understanding the function and dynamics of any particular network (e.g. [68]). Such findings are emphasized by in silico evolution studies, which find network dynamics rather than structure to be the determining and conserved features of the cellular networks underlying complex developmental phenotypes (e.g. [17, 68–71]).

The capacity of ESB to reformulate or produce entirely new questions about evolutionary dynamics in light of systems biology is clearly illustrated by the study of noise in biological systems. Noisy dynamics arising at the cellular level can manifest as phenotypic variance at the population level [72]. This phenomenon raises questions about how noise might

affect evolutionary dynamics and which evolutionary conditions favor the selection of the underlying mechanisms. The latter question arises from the fact that noise levels themselves are subject to tuning via evolutionary processes such as point mutations in regulatory regions [73]. This interplay between noise, evolutionary dynamics, and systems behavior sets the stage perfectly for an integrative ESB approach. For example, there are indications from genomic analyses that certain functional gene classes exhibit significant increases in expression noise [74]. This correlation can be explained by system-dynamic and population-genetic models that explore the types of genotype-phenotype maps that could lead to the selection of increased expression noise [74, 75]. Further ESB studies are exploring the phenotypic consequences of noise and its impact on the evolutionary dynamics of metabolic networks [76].

In regulatory networks, noise combined with nonlinear gene-regulatory dynamics can lead to the bimodal distribution of phenotypes (i.e. phenotype switching) in otherwise isogenic populations [77, 78]. This is vividly illustrated in the case of antibiotic resistance. In a genetically identical population of bacteria, high doses of antibiotics kill most but not all organisms. Upon re-culturing, the surviving bacteria give rise to a population that is natively affected by antibiotics [78]. These resistant cells are thus phenotypic rather than genetic variants. Similar observations are made in the case of sporulation and DNA-uptake pathways, where pathway activity shows a significant variance across the population [77, 79]. While systems biology is still discovering the mechanistic bases of phenotypic variance [80], ESB provides a conceptual context in which to evaluate these findings [81–83]. In particular, a recent in silico study suggests that phenotypic switching in a gene-regulatory system can emerge as a byproduct of the evolution of elevated nonlinearity, which is itself selected for its impact on noise and evolvability [84]. Being able to examine phenotypic switching in laboratory-based experimental evolution [85] allows hypotheses about the evolution of noise to be tested (e.g. regarding the mutational steps toward the molecular mechanisms of noise).

Broadening the scope of ESB

ESB is developing rapidly in several directions. While some of these initiatives advance established research areas, others set up new research agendas on the basis of methodological innovations.

Expanding evolutionary and ecological frameworks

The cross-fertilization of different approaches in ESB is rebalancing the way in which evolution is understood to shape cellular networks. Rather than maintaining a division between adaptive and non-adaptive explanations, as often occurs in evolutionary research, ESB analyses are showing how both aspects are needed to explain evolved systems. Neutral processes can be understood to be co-dynamical with adaptive processes in network evolution [86]. For instance, in the case of modularity, several studies indicate both adaptive and neutral scenarios to be of importance for modularity's emergence. These investigations indicate the environmental and evolutionary scenarios that can lead to the evolution of modularity: fluctuating selection in environments composed of modular tasks [21, 87], genetic drift [22], and stabilizing selection combined with selection for novel functions [88–90].

Although it will always be difficult to discern which plausible scenario has produced particular network features in specific organisms, in silico evolution and system-level mathematical models allow assessments of ecological and evolutionary conditions (such as fluctuating environments and population size) to be linked to network properties (and vice-versa). Rigorous evaluation of evolutionary scenarios can thereby incorporate comparative analyses of global network structures [15, 16, 54, 91], and together they can comprise system-based "reverse ecology" [92]. Reverse ecology in ESB aims to infer the historical ecologies of organisms from features of present-day cellular networks (rather than focusing on genetic loci as population genomics "reverse ecology" does). ESB aspires toward an increased ability to predict the future evolution of organisms (particularly single-celled organisms) in response to novel selective pressures and environments. This outcome is already within grasp in the case of metabolic networks [6].

Encompassing intercellular network evolution

Selective pressures imposed by the abiotic environment on biological systems are only one component of ongoing evolutionary dynamics, which are always entwined with interspecies interactions (Fig. 1). While emerging systems-biological approaches are now beginning to scrutinize these interactions at the level of their underlying intracellular networks, one strand of ESB aims to understand more fully the effects of species interactions on network evolution. For example, the increasingly detailed characterization of parasite manipulation of host networks, at the level of single proteins and whole networks [93], has produced hypotheses about how such effects shape host network evolution [94]. Recent efforts to extend the application of FBA to species interactions do this by combining computational analysis with experiments [29, 30, 95]. These applications permit an examination of the environmental conditions and co-evolutionary dynamics that enable species to co-exist in syntrophy.

Important open questions in this area of research include how to understand the evolution and dynamics of diverse microbial communities in the context of their intercellular networks, and how the network-level impact of host-parasite interactions can be captured more effectively. Taking an ESB approach, combined with the increasing ability to probe species interactions experimentally and genomically in natural contexts such as soil microcosms [96], expands the scope of how microbial communities are investigated. In addition, ESB studies that analyse cellular networks in light of social evolution can assess the impact of competitive or cooperative interactions on these networks, and in turn predict how they enable future competitive and cooperative dynamics [31].

Re-engineering networks in light of evolution

A major future contributor to ESB is likely to be synthetic biology, which introduces an engineering approach to evolved systems. By accelerating the tinkering to which evolution has subjected biological systems, and incorporating mathematically redesigned features, synthetic biology can intervene in evolutionary processes and outcomes at the systems level. For example, synthetic circuits could be constructed to test hypotheses about intermediary states in the evolution of specific cellular networks, or to measure the effect of mutations in different genetic backgrounds.

Combining synthetic biology with experimental evolution could lead to an even greater capacity to test evolutionary hypotheses. Exposing synthetic circuits to further evolution under a variety of conditions (e.g. fluctuating versus stable environments), or developing high-throughput techniques to generate a diversity of circuits (e.g. [97]), allows the examination of both the effects of specific network features on future evolution and the role of environmental conditions on the outcome of network evolution. While recent experimental evolution has produced detailed characterizations of the molecular steps toward specific evolutionary trajectories and outcomes in natural systems (e.g. [98, 28]), combining experiments with synthetic biology and some of the theoretical approaches from ESB (e.g. dynamical network models, flux balance analysis) should increase research scope and depth.

Integrating ESB and biomedicine

The improved ability to specify how and why networks have evolved will be particularly beneficial when the evolutionary outcomes relate to human health, as demonstrated by the investigation of the evolution of antibiotic resistance (e.g. [28]). So far, evolutionary medicine [99] and systems or network medicine [100] have interacted very little, and both are just nascent fields or even mere ambitions in the minds of some researchers. But it is very probable that many disease and treatment systems, from cancer to antibiotic resistance, are unlikely to be comprehended effectively or made tractable for intervention without being modeled as dynamic evolving systems [101, 102]. Using mathematical modeling and synthetic-biological constructions to predict, for example, how antibiotic resistance will evolve in relation to network vulnerabilities to specific molecular interactions, would have sustainable therapeutic effects [103]. Drug development, foundering in many traditional approaches (especially when restricted to the selection of target molecules), may thus yield to a synthesis of evolutionary and systems biological approaches because it steers drug development back to a focus on interactions within an evolving physiological context [104].

Synthesizing explanation and prediction

We mentioned above that a central aim of ESB is to produce mechanistic and evolutionary understanding of genotype-phenotype relations. Mechanistic explanations are causal accounts of how certain outcomes occur in specific circumstances. In the ESB case, the focus is on how particular genotypic features produce certain phenotypic properties. Historically, molecular biology has offered qualitative mechanistic explanations. One aspect of systems biology has worked to make these explanations quantitatively mechanistic rather than primarily descriptive-predictive. Dynamic mathematical modeling and time-series data are crucial to such explanations. ESB goes one step further and attempts to embed quantitative mechanistic explanation (which in Ernst Mayr's terminology [105] is proximate explanation), within ultimate or evolutionary explanation.

Prediction has sometimes been suggested to be an achievement beyond the capacity of evolutionary biology (e.g. [105, 106, 107]). Although evolutionary biology is accepted as explanatory, it has been placed at the ultimate end of the spectrum of explanation rather than the proximate, and the field has often been charged with untestability and panadaptationism. Systems biology on the other hand has focused largely on description and prediction (i.e. characterizing systems and predicting fluctuations of that system under specific perturbation regimes). The aim to achieve both ultimate and proximate explanation, combined with predictiveness, is a major motivation of ESB. This does not mean its practitioners expect to find universal laws. But they do expect to find regularities that hold across a wide range of networks and organisms as a result of common evolutionary processes, and there is good reason to think that finding these will help explain and predict system behavior. This is because evolutionary processes, despite being both stochastic and contingent, can leave stable traces of their impact on system architecture and function. These traces can then be linked back to evolutionary and ecological conditions. When this historical knowledge is combined with detailed mechanistic explanation, evolutionary trajectories can be projected forward and sought in silico or in vitro. Evolutionary predictions will never be easy to formulate and confirm, but ESB offers a platform for developing the predictive aspects of evolutionary biology and in the process, refining how evolutionary prediction is understood (because of the necessity of taking into account contingency and uncertainty).

Some of the basis for predictive ESB is already established in metabolic network research [6]. These predictions currently focus on detecting the effects of mutations on metabolic function, identifying environments from the structure and function of metabolic networks, and establishing causal connections between structural features of metabolic networks and ecological conditions. Extending this basic predictive capacity will mean going beyond metabolic networks to other cellular interactions, by building predictive models based on explanations of the origins and diversity of networks as they respond temporally to complex conditions.

Conclusions and outlook

We have presented ESB as an emerging but loose synthesis of system approaches to biological phenomena

interpreted within an overarching evolutionary framework. However, there are numerous disciplinary and training barriers to be overcome to enable researchers to contribute to ESB. From the perspective of many quantitative and modeling-oriented biologists, evolutionary research consists of historical narratives that are not "hard" science. And from the evolutionary biologists' perspective, their own field already has sufficiently rigorous methods at its disposal. ESB is in one sense a demonstration of the usefulness of thinking outside these disciplinary boxes – not to make researchers switch disciplinary allegiance but to enhance their own disciplinary achievements. For this reason, we advocate a more integrative systems biology curriculum – able to encompass evolutionary aspects of systems – at whatever educational levels system biology is taught.

Although we do not anticipate all evolutionary biologists becoming ESB advocates or all systems biologists signing up for evolutionary analyses, we do foresee that integrative ESB efforts will have increasing payoffs for difficult research questions. Describing, explaining, and predicting genotype-phenotype relationships will obviously need a great deal of experimental work to ground studies based on dynamic models and evolutionary simulations. However, if there is one thing molecular biology has learned in the last decade it is that integrative combinations of modeling and experimental approaches are crucial to more predictive and explanatory understandings of systems. All the caveats about modeling uncertainties, data shortcomings and experimental intractability that apply to systems biology apply to ESB. But as this review illustrates, many steps forward have been made and this is in part because the field is characterized more by research-driven inquiry than grand ambitions to create a new disciplinary entity.

We must emphasize that we are neither claiming that ESB is a panacea to disciplinary shortcomings, nor that it has all the right questions and answers. Instead, we are drawing attention to the broad phenomenon of ESB, rather than a specific strand here or there, because of the field's general implications for the integration of certain modes of explanation and prediction. Even for scientists not interested in evolutionary or systems biology, these generalities are issues of scientific relevance. It may be the case that ESB becomes so broad in its problem orientation and methodology that it cannot be demarcated even as a field, let alone a discipline: it could simply become a general approach to the study of evolving biological systems. This, we believe, would be more valuable than if ESB were to form a tightly bounded field with disciplinary ambitions.

Acknowledgements

We thank our referees and editor for very helpful comments, and several colleagues for comments on a much earlier version of this paper. MAO acknowledges support from the Australian Research Council and University of Sydney in the form of a Future Fellowship. We are grateful to Arno Steinacher for creating the two figures.

References

1. **Medina M**. 2005. Genomes, phylogeny, and evolutionary systems biology. *Proc Natl Acad Sci USA* **102**: 6630–5.
2. **Koonin EV, Wolf YI**. 2006. Evolutionary systems biology: links between gene evolution and function. *Curr Opin Biotechnol* **17**: 481–7.
3. **Dean AM, Thornton JW**. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**: 675–88.
4. **O'Malley MA**. 2012. Evolutionary systems biology: historical and philosophical perspectives on an emerging synthesis. In Soyer OS, ed; *Evolutionary Systems Biology*. NY: Springer. p. 1–28.
5. **Soyer OS**. ed; 2012. *Evolutionary Systems Biology*. NY: Springer.
6. **Papp B, Notebaart RA, Pál C**. 2011. Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet* **12**: 591–602.
7. **Loewe L**. 2009. A framework for evolutionary systems biology. *BMC Syst Biol* **3**: 27.
8. **Morange M**. 1998. *A History of Molecular Biology*. Cambridge MA: Harvard University Press.
9. **Boogerd FC, Bruggeman FJ, Hofmeyr J-HS, Westerhoff HV**. eds; 2007. *Systems Biology: Philosophical Foundations*. Amsterdam: Elsevier.
10. **Alon U, Surette MG, Barkai N, Leibler S**. 1999. Robustness in bacterial chemotaxis. *Nature* **397**: 168–71.
11. **Klipp E, Nordlander B, Krüger R, Gennemark P**, et al. 2005. Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* **23**: 975–82.
12. **Jaeger J, Surkova S, Blagov M, Janssens H**, et al. 2004. Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**: 368–71.
13. **Gao F, Davidson EH**. 2008. Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proc Natl Acad Sci USA* **105**: 6091–6.
14. **Hoyos E, Kim K, Milloz J, Barkoulas M**, et al. 2011. Quantitative variation in autocrine signaling and pathway crosstalk in the *Caenorhabditis* vulval network. *Curr Biol* **21**: 527–38.
15. **Nikolaou E, Agrafioti I, Stumpf M, Quinn J**, et al. 2009. Phylogenetic diversity of stress signalling pathways in fungi. *BMC Evol Biol* **9**: 44.
16. **Wuichet K, Zhulin IB**. 2010. Origins and diversification of a complex signal transduction system in prokaryotes. *Sci Signal* **3**: ra50.
17. **François P, Hakim V, Siggia ED**. 2007. Deriving structure from evolution: metazoan segmentation. *Mol Syst Biol* **3**: 154.
18. **Goldstein RA, Soyer OS**. 2008. Evolution of taxis responses in virtual bacteria: non-adaptive dynamics. *PLoS Comput Biol* **4**: e1000084.
19. **Thorne TW, Ho H-L, Huvet M, Haynes K**, et al. 2011. Prediction of putative protein interactions through evolutionary analysis of osmotic stress response in the model yeast *Saccharomyces cerevisae*. *Fungal Genet Biol* **48**: 504–11.
20. **Alon U**. 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**: 450–61.
21. **Kashtan N, Alon U**. 2005. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci USA* **102**: 13773–8.
22. **Lynch M**. 2007. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* **8**: 803–13.
23. **Wagner A**. 2008. Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet* **9**: 965–74.
24. **Giaever G, Chu AM, Ni L, Connelly C**, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–91.
25. **Orr HA**. 2005. The genetic theory of adaptation: a brief history. *Nat Rev Genet* **6**: 119–27.
26. **Akman OE, Rand DA, Brown PE, Millar AJ**. 2010. Robustness from flexibility in the fungal circadian clock. *BMC Syst Biol* **4**: 88.
27. **Feist AM, Herrgård MJ, Thiele I, Reed JL**, et al. 2008. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7**: 129–43.
28. **Toprak E, Veres A, Michel J-B, Chait R**, et al. 2012. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet* **44**: 101–5.
29. **Klitgord N, Segrè D**. 2010. Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol* **6**: e1001002.
30. **Bordbar A, Lewis NE, Schellenberger J, Palsson BØ**, et al. 2010. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol Syst Biol* **6**: 422.
31. **Foster KR**. 2011. The sociobiology of molecular systems. *Nat Rev Genet* **12**: 193–203.
32. **Emmerling M, Dauner M, Ponti A, Fiaux J**, et al. 2002. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J Bacteriol* **184**: 152–64.

33. **de Visser JA, Hermisson J, Wagner GP, Meyers LA**, et al. 2003. Perspective: evolution and detection of genetic robustness. *Evolution* **57**: 1959–72.
34. **Wagner A**. 2000. The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* **154**: 1389–401.
35. **Krakauer DC, Plotkin JB**. 2002. Redundancy, antiredundancy, and the robustness of genomes. *Proc Natl Acad Sci USA* **99**: 1405–9.
36. **Bergman A, Siegal ML**. 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**: 549–52.
37. **Azevedo RBR, Lohaus R, Srinivasan S, Dang KK**, et al. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**: 87–90.
38. **Soyer OS, Pfeiffer T**. 2010. Evolution under fluctuating environments explains observed robustness in metabolic networks. *PLoS Comput Biol* **6**: e1000907.
39. **Harrison R, Papp B, Pál C, Oliver SG**, et al. 2007. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci USA* **104**: 2307–12.
40. **Siegal ML, Bergman A**. 2002. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci USA* **99**: 10528–32.
41. **Wang Z, Zhang J**. 2009. Abundant indispensable redundancies in cellular metabolic networks. *Genome Biol Evol* **1**: 23–33.
42. **Phillips PC**. 2008. Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**: 855–67.
43. **Wagner GP, Zhang J**. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* **12**: 204–13.
44. **Segrè D, Deluna A, Church GM, Kishony R**. 2005. Modular epistasis in yeast metabolism. *Nat Genet* **37**: 77–83.
45. **Snitkin ES, Segrè D**. 2011. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genet* **7**: e1001294.
46. **Gutiérrez J**. 2009. A developmental systems perspective on epistasis: computational exploration of mutational interactions in model developmental regulatory networks. *PLoS ONE* **4**: e6823.
47. **Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg O**. 2007. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* **175**: 411–20.
48. **Landry CA, Rifkin SA**. 2012. The genotype-phenotype maps of systems biology and quantitative genetics: distinct but complementary. In Soyer OS, ed; *Evolutionary Systems Biology*. NY: Springer. p. 371–98.
49. **Moore JH, Williams SM**. 2005. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *BioEssays* **27**: 637–46.
50. **Tyler AL, Asselbergs FW, Williams SM, Moore JH**. 2009. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *BioEssays* **31**: 220–7.
51. **Wang Z, Liao BY, Zhang J**. 2010. Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci USA* **107**: 18034–9.
52. **Hill WG, Zhang ZS**. 2012. On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. *Genetics* **190**: 1131–7.
53. **Knight CG, Zitzmann N, Prabhakar S, Antrobus R**, et al. 2006. Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. *Nat Genet* **38**: 1015–22.
54. **Huvet M, Toni T, Sheng X, Thorne T**, et al. 2011. The evolution of the phage shock protein response system: interplay between protein function, genomic organization, and system function. *Mol Biol Evol* **28**: 1141–55.
55. **Beltrao P, Trinidad JC, Fiedler D, Roguev A**, et al. 2009. Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol* **7**: e1000134.
56. **Madan Babu M, Teichmann SA, Aravind L**. 2006. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358**: 614–33.
57. **Tan K, Shlomi T, Feizi H, Ideker T**, et al. 2007. Transcriptional regulation of protein complexes within and across species. *Proc Natl Acad Sci USA* **104**: 1283–8.
58. **Carroll SB**. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.
59. **Peter IS, Davidson EH**. 2011. Evolution of gene regulatory networks controlling body plan development. *Cell* **144**: 970–85.
60. **Erwin DH, Davidson EH**. 2009. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* **10**: 141–8.
61. **Alonso CR, Wilkins AS**. 2005. The molecular elements that underlie developmental evolution. *Nat Rev Genet* **6**: 709–15.
62. **Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S**, et al. 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**: 594–7.
63. **Chen K, Rajewsky N**. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**: 93–103.
64. **Jaeger J, Irons D, Monk N**. 2012. The inheritance of process: a dynamical systems approach. *J Exp Zool (Mol Dev Evol)* **318B**: 591–612.
65. **Davidson EH**. 2010. Emerging properties of animal gene regulatory networks. *Nature* **468**: 911–20.
66. **Manu, Surkova S, Spirov AV, Gursky VV**, et al. 2009. Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. *PLoS Biol* **7**: e1000049.
67. **Salazar-Ciudad I, Jernvall J**. 2010. A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**: 583–6.
68. **Siegal ML, Promislow DE, Bergman A**. 2007. Functional and evolutionary inference in gene networks: does topology matter? *Genetica* **129**: 83–103.
69. **Fujimoto K, Ishihara S, Kaneko K**. 2008. Network evolution of body plans. *PLoS ONE* **3**: e2772.
70. **ten Tusscher KH, Hogeweg P**. 2011. Evolution of networks for body plan patterning; interplay of modularity, robustness and evolvability. *PLoS Comput Biol* **7**: e1002208.
71. **François P, Siggia ED**. 2010. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* **137**: 2385–95.
72. **Eldar A, Elowitz MB**. 2010. Functional roles for noise in genetic circuits. *Nature* **467**: 167–73.
73. **Blake WJ, Balázsi G, Kohanski MA, Issacs FJ**, et al. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* **24**: 853–65.
74. **Zhang Z, Qian W, Zhang J**. 2009. Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol* **5**: 299.
75. **Tănase-Nicola S, ten Wolde PR**. 2008. Regulatory control and the costs and benefits of biochemical noise. *PLoS Comput Biol* **4**: e1000125.
76. **Wang Z, Zhang J**. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci USA* **108**: E67–76.
77. **Maamar H, Raj A, Dubnau D**. 2007. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**: 526–7.
78. **Balaban NQ, Merrin J, Chait R, Kowalik L**, et al. 2004. Bacterial persistence as a phenotypic switch. *Science* **305**: 1622–5.
79. **Veening J-W, Stewart EJ, Berngruber TW, Taddei F**, et al. 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci USA* **105**: 4393–8.
80. **Levy SF, Ziv N, Siegal ML**. 2012. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol* **10**: e1001325.
81. **Levy SF, Siegal ML**. 2012. The robustness continuum. In Soyer OS, ed; *Evolutionary Systems Biology*. NY: Springer. p. 431–452.
82. **Kussell E, Kishony R, Balaban NQ, Leibler S**. 2005. Bacterial persistence: a model of survival in changing environments. *Genetics* **169**: 1807–14.
83. **Salathé M, Van Cleve J, Feldman MW**. 2009. Evolution of stochastic switching rates in asymmetric fitness landscapes. *Genetics* **182**: 1159–64.
84. **Kuwahara H, Soyer OS**. 2012. Bistability in feedback circuits as a byproduct of evolution of evolvability. *Mol Syst Biol* **8**: 564.
85. **Beaumont HJ, Gallie J, Kost C, Ferguson GC**, et al. 2009. Experimental evolution of bet hedging. *Nature* **462**: 90–3.
86. **Wagner A**. 2011. *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*. NY: Oxford University Press.
87. **Parter M, Kashtan N, Alon U**. 2008. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput Biol* **4**: e1000206.
88. **Soyer OS**. 2007. Emergence and maintenance of functional modules in signaling pathways. *BMC Evol Biol* **7**: 205.
89. **Hintze A, Adami C**. 2008. Evolution of complex modular biological networks. *PLoS Comput Biol* **4**: e23.
90. **Espinosa-Soto C, Wagner A**. 2010. Specialization can drive the evolution of modularity. *PLoS Comput Biol* **6**: e1000719.
91. **Borenstein E, Kupiec M, Feldman MW, Ruppin E**. 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci USA* **105**: 14482–7.

92. **Levy R, Borenstein E**. 2012. Reverse ecology: from systems to environments and back. In Soyer OS, ed; *Evolutionary Systems Biology*. ed; NY: Springer. p. 329–45.
93. **Bhavsar AP, Guttman JA, Finlay BB**. 2007. Manipulation of host-cell pathways by bacterial pathogens. *Nature* **449**: 827–34.
94. **Salathé M, Soyer OS**. 2008. Parasites lead to evolution of robustness against gene loss in host signaling networks. *Mol Syst Biol* **4**: 202.
95. **Hillesland KL, Stahl DA**. 2010. Rapid evolution of stability and productivity at the origin of a microbial mutualism. *Proc Natl Acad Sci USA* **107**: 2124–9.
96. **Gómez P, Buckling A**. 2011. Bacteria-phage antagonistic coevolution in soil. *Science* **332**: 106–9.
97. **Wang HH, Issacs FJ, Carr PA, Sun ZZ**, et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**: 894–8.
98. **Woods RJ, Barrick JE, Cooper TF, Shrestha U**, et al. 2011. Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**: 1433–6.
99. **Williams GC, Nesse RM**. 1991. The dawn of Darwinian medicine. *Q Rev Biol* **66**: 1–22.
100. **Barabási AL, Gulbahce N, Loscalzo J**. 2011. Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**: 56–68.
101. **Gatenby RA, Gillies RJ, Brown JS**. 2010. Evolutionary dynamics of cancer prevention. *Nat Rev Cancer* **10**: 526–7.
102. **Baquero F, Coque TM, de la Cruz F**. 2011. Ecology and evolution as targets: the need for novel eco-evo drugs and strategies to fight antibiotic resistance. *Antimicrob Agents Chemother* **55**: 3649–60.
103. **Palmer AC, Kishony R**. 2013. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet* **14**: 243–8.
104. **Pujol A, Mosca R, Farrés J, Aloy P**. 2010. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci* **31**: 115–23.
105. **Mayr E**. 1961. Cause and effect in biology. *Science* **134**: 1501–6.
106. **Peters RH**. 1976. Tautology in evolution and ecology. *Am Nat* **110**: 1–12.
107. **Gould SJ**. 1990. *Wonderful Life: The Burgess Shale and the Nature of History*. London: Hutchinson Radius.

Skinner et al. BMC Genomics 2014, 15:377
http://www.biomedcentral.com/1471-2164/15/377

BMC Genomics

RESEARCH ARTICLE Open Access

Gene bionetworks involved in the epigenetic transgenerational inheritance of altered mate preference: environmental epigenetics and evolutionary biology

Michael K Skinner[1*], Marina I Savenkova[1], Bin Zhang[2], Andrea C Gore[3] and David Crews[4]

Abstract

Background: Mate preference behavior is an essential first step in sexual selection and is a critical determinant in evolutionary biology. Previously an environmental compound (the fungicide vinclozolin) was found to promote the epigenetic transgenerational inheritance of an altered sperm epigenome and modified mate preference characteristics for three generations after exposure of a gestating female.

Results: The current study investigated gene networks involved in various regions of the brain that correlated with the altered mate preference behavior in the male and female. Statistically significant correlations of gene clusters and modules were identified to associate with specific mate preference behaviors. This novel systems biology approach identified gene networks (bionetworks) involved in sex-specific mate preference behavior. Observations demonstrate the ability of environmental factors to promote the epigenetic transgenerational inheritance of this altered evolutionary biology determinant.

Conclusions: Combined observations elucidate the potential molecular control of mate preference behavior and suggests environmental epigenetics can have a role in evolutionary biology.

Keywords: Epigenetics, Brain, Networks, Evolution, Behavior

Background

The current molecular paradigm for neo-Darwinian evolutionary biology is that random DNA sequence mutations, gene flow, and genetic drift promote phenotype variation that allows an adaptation event to facilitate natural selection [1]. Although environment has an important role in natural selection, environmental factors generally do not have the capacity to alter DNA sequence or mutation rates. A small group of compounds can act as mutagens, but the vast majority of nutritionally derived and environmental toxicants do not alter DNA sequence [2]. The current genetic paradigm does not completely explain many observations such as rapid evolutionary events, environmental impacts on evolution, and the low frequency of the occurrence of useful mutations [3,4]. The realization that epigenetics provides an additional molecular mechanism for the environment to influence genome activity and biology has suggested a potential role for environmental epigenetics in evolutionary biology [5-11]. Charles Darwin recognized sexual selection as one of two determinants in evolutionary biology, the other being natural selection [12]. The physical attributes and courtship rituals involved in mate preference are essential for reproductive fitness and propagation of a species. The current study examines how environmental factors can promote an epigenetic event to promote an alteration in mate preference behavior.

Previously we demonstrated that exposure of a gestating female rat to an environmental compound during fetal gonadal sex determination promoted epigenetic reprogramming of the male germline [13-15]. These reprogrammed differential DNA methylation regions (DMR) in the sperm epigenome have recently been shown to be induced by a

* Correspondence: skinner@wsu.edu
[1]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA
Full list of author information is available at the end of the article

variety of different environmental toxicants with exposure specific DMR [16,17]. The initial environmental compound used was the commonly used fungicide vinclozolin which is an anti-androgenic endocrine disruptor [18]. The primordial germ cells during migration down the genital ridge undergo a DNA methylation erasure that then upon gonadal sex determination the DNA re-methylation is initiated in a sex-specific manner [19]. Environmental exposures during this developmental stage modifies the epigenetic programming of the male germline that becomes re-programmed (imprinted-like) and promotes a transgenerational phenotypic variation and adult onset disease state in subsequent generations [13-16]. The epigenetic transgenerational inheritance of adult onset disease (i.e. after one year of age) in males includes infertility, prostate disease, kidney disease, immune abnormalities and spermatogenic defects [20,21], and in females includes mammary tumor development, kidney disease, reproductive tissue abnormalities and pregnancy abnormalities [22]. This germline mediated epigenetic transgenerational inheritance of adult onset disease is mediated in part through alterations in the sperm epigenome [13,14]. Since the germline establishes the base line epigenome of the organism, all tissues in both the female and male progeny including the brain appear to have altered tissue specific transgenerational epigenomes, transcriptomes and phenotypes [15,21,23-25].

Investigation of the epigenetic transgenerational inheritance of altered brain genome activity and behaviors previously demonstrated anxiety-like behavior increased in females and decreased in males, which correlate to alterations in specific brain region transcriptomes [23]. Altered stress responses are also detected in the transgenerational exposure lineage animals [26]. Interestingly, previous analysis of F3 generation control and vinclozolin lineage female and male rats (i.e. prior to the onset of disease) demonstrated an alteration in mate preference behavior [27]. The female rats, independent of control or vinclozolin lineage, prefer control lineage males if given a choice. This behavioral decision raises the possibility of an epigenetic contribution to mate preference and sexual selection. The current study was designed to directly correlate the altered mate preference behavior with gene networks in specific brain regions in both the females and males. Observations elucidate the potential molecular control of mate preference behavior and demonstrates environmental factors have the capacity to promote the epigenetic transgenerational inheritance of altered mate preference.

Systems biology analysis has allowed biological phenomena such as mate preference to be considered from the molecular to physiological level. The gene bionetwork [28] analysis previously developed to investigate the molecular basis of disease [29] was used in the current study. This approach has been used to identify gene networks associated with disease, such as obesity and diabetes [30]. Recently, we have used this bionetwork analysis to study a normal developmental process of primordial follicle development in the ovary [31,32]. The gene networks identified were found to contain growth factors that are known to regulate the developmental process [31,32]. These bionetwork analyses use a large number of microarray transcriptome analyses under different perturbations to identify gene clusters and modules that are coordinately regulated [33,34]. The gene networks observed identify the genes with the highest level of integration and connection (i.e. connectivity) that associate with the phenotype [29,33-35]. This genomic approach was used in the current study to identify the gene bionetworks in various brain regions associated with mate preference.

Observations demonstrate an environmental compound exposure can induce an epigenetic reprogramming of the germline that promotes epigenetic transgenerational inheritance of altered mate preference behavior. Although no direct epigenetic modifications in the brain were examined, the environmentally induced epigenetic transgenerational model used indicates epigenetics can be involved in the induction of the altered behavioral phenotypes. Sex-specific effects were observed in both the male and female brain transcriptome and behavior correlations. The gene networks in specific brain regions that statistically correlate with various mate preference behaviors provides insight into this environmentally modified transgenerational behavior. This systems biology approach has elucidated novel mechanisms to be considered in mate preference biology.

Results

The experimental design involved the development of transgenerational control and vinclozolin lineage animals for a mate preference behavioral analysis [27]. Subsequently, a transcriptome analysis was performed on 6 different brain regions from adult male and female F3 generation Sprague Dawley rats. These brain regions have previously been shown to be associated with mate preference behavior [27,36]. The transcriptome alterations were statistically correlated with changes in mate preference behaviors. As previously described [13,37], F0 generation gestating females were transiently exposed daily to vehicle control DMSO or vinclozolin from embryonic day 8–14 (E8-14) during fetal gonadal sex determination. The F1 generation offspring were bred at 90 days of age to generate F2 generation control and vinclozolin lineage progeny and then F2 generation animals were bred to generate the F3 generation control and vinclozolin lineage animals [13]. No sibling or cousin breeding was used to avoid any inbreeding artifacts. The F3 generation control and vinclozolin male and female rats were analyzed at 3–4 months

of age for mate preference behaviors, as previously described [27]. This is an age when no major adult onset disease has been detected or is anticipated [20]. Later at 11 month of age, animals were sacrificed and specific brain regions isolated and RNA collected for microarray transcriptome analysis. The differentially regulated gene sets ("Signature lists") for each brain region were identified. Subsequently a bioinformatics bionetwork analysis [23,31] was used to correlate gene modules and networks with mate preference behaviors observed (Figure 1).

Females of both control and vinclozolin lineages preferred control lineage males over vinclozolin lineage males [27]. The mate preference behaviors are described in detail in the Methods [27] (Additional file 1: Movie S1). The specific behavioral components associated with mate preference analysis include the following: "Wire Mesh" involved the experimental animal investigating the stimulus animals directly through the Wire Mesh; "Facial Investigation" entailed the actual nose-to-nose contact; "Plexiglas" refers to the experimental animal investigating the area immediately bordering the Wire Mesh that separated the experimental animal from the stimulus animal; "Walking" refers to general investigation of the central testing chamber as measured by undirected walking and sniffing; and "Still" in which the animal was stationary with minimal head movement. Additional file 2: Table S1A presents the values for each behavioral component associated with individual animals. After completion of the behavioral analysis the animals were sacrificed at 11 months of age and 6 different brain regions dissected including the amygdala (Amy), hippocampus (Hipp), olfactory bulb (OlfB), cingulate cortex (CngCtx), entorhinal cortex (EnCtx), and preoptic area-anterior hypothalamus (POAH). The procedure to isolate the brain regions is described in the Materials and Methods. The isolated tissue was immediately placed in Trizol reagent, frozen and stored. RNA was prepared for microarray transcriptome analysis from each animals brain regions independently.

For the microarray analysis each F3 generation control and vinclozolin lineage male and female animal had six different brain regions analyzed which totaled 134 different microarrays. The microarray data were pre-processed and demonstrated two abnormal arrays that were omitted for further analysis (Additional file 3: Figure S1B). Batch effect corrections were made for RNA preparation date and array scan date with no major batch effects detected. The array data were then processed as previously described [31] to identify the differentially expressed gene sets for each brain region (Table 1). The differentially expressed genes in the Signature lists required a greater than 1.2 fold change in expression and all changes in expression were statistically significant with $p < 0.05$, as described in the Methods. Since a 20% alteration in gene expression for many genes, such as transcription factors, can have dramatic cellular and biological responses [26,32], a more stringent cut off (e.g. 2×) was not used in the current study. In the current study the primary focus was on the coexpression patterns of the differentially expressed genes through the co-expression network analysis.

The number of control lineage versus vinclozolin lineage differentially expressed genes in the Signature lists ranged from 43 to 803 with both up-regulated and down-regulated genes (Table 1). The total number of control versus vinclozolin lineage differentially expressed genes for all brain regions combined was 1833 for females and 1693 for males. A list of all the genes separated by brain region, sex and functional gene categories is presented in Additional file 4: Table S2 A-I. The overlap and differences between the

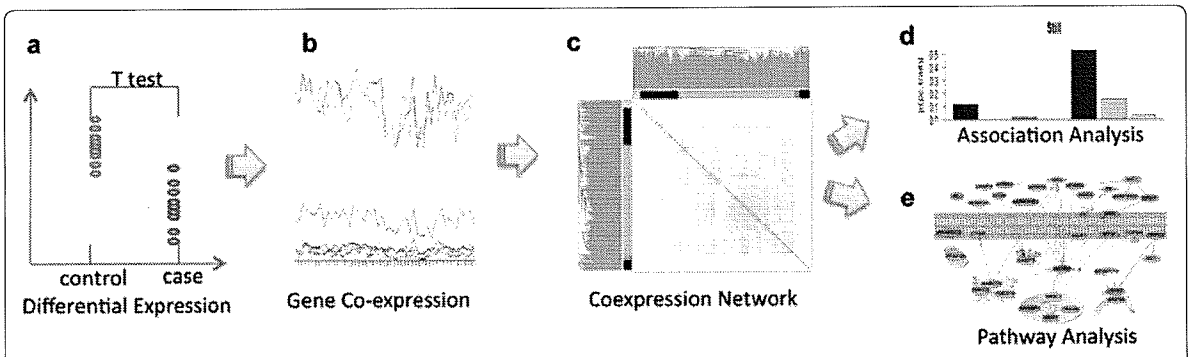

Figure 1 A flowchart of the analyses carried out in the study. (a) Differentially expressed (DE) genes are identified by the standard *t*-test. **(b)** Correlations between the expression profiles of the DE genes are calculated to quantify the level of coexpression. **(c)** Coexpression network analysis is performed to identify coexpressed gene modules. A matrix of correlations between gene expression profiles is first transformed through a power function into an adjacency matrix that is further transformed into a topological overlap matrix (TOM). Modules (represented by color bars) comprised of highly coexpressed genes are the identified using a dynamic cut-tree algorithm. **(d)** Association between gene modules and phenotypic traits is then accessed. **(e)** Pathway analysis is then performed on the gene modules of interest to derive regulatory networks for mechanism discovery.

Table 1 Differentially expressed Signature genes and their overlap with modules generated in combined network

Sex-region	Signature lists	Over-expressed	Under-expressed	Combined networks modules										Separate network modules						
				Number of modules	Turquoise	Blue	Brown	Yellow	Green	Red	Black	Pink	Magenta	Number of modules	Turquoise	Blue	Brown	Yellow	Green	Red
Female regions	1833	939	894	4	1090	283	259	104							# genes in module					
				# genes overlapped between module & Signature list																
F-Amy	139	38	101		81	19	15	0						3	71*	49	17			
F-CngCTX	803	481	322		640	16	57	83						4	444	183	82	74		
F-EnCTX	433	279	154		56	191	128	37						3	369*	35	10*			
F-Hipp	70	40	30		18	17	12	0						1	70					
F-OlfB	748	221	527		598	59	69	3						4	416	305	11	10*		
F-POAH	56	24	32		16	18	11	1						1	56*					
Male regions	1693	638	1055	9	505	287	222	155	88	66	50	40	36			# genes in module				
				# genes overlapped between module & Signature list																
M-Amy	175	105	70		35	19	15	3	25	8	10	8	13	2	160*	10*				
M-CngCTX	785	189	596		354	79	193	39	56	8	0	5	0	1	780*					
M-EnCTX	385	210	175		87	133	26	8	13	3	0	5	22	1	378*					
M-Hipp	151	30	121		13	27	0	11	2	41	9	21	0	2	133*	13*				
M-OlfB	356	278	78		71	33	0	114	16	9	47	7	9	6	231	65*	20*	12*	11*	11
M-POAH	43	19	24		3	13	1	4	1	0	1	7	0	1	43*					

*- modules that showed statistically significant correlation with behavior.

Signature lists of each brain region for male and female is shown in Figure 2. The majority of genes were distinct to the different lists in a comparison of the brain regions. The one exception was an overlap between the cingulate cortex (CngCtx) and olfactory bulb (OlfB) in the female. Therefore, each brain region Signature list was distinct from each other and between the sexes.

Analysis of the cellular pathways and processes associated with the gene Signature lists for each brain region is shown in Figure 2 and Additional file 5. The top 36 pathways with the greatest combined number of genes associated are shown in Additional file 5. An extended list of pathways and processes with the associated genes from the different modules and tissues is presented in Additional file 6: Table S3. Several of the most highly represented pathways in the male and female were the MAPK signaling pathway, olfactory transduction, neuroactive ligand-receptor interactions and axon guidance. The Signature list genes distributed relatively evenly across the different pathways with no major over-representation identified. Most major cellular processes and pathways were represented with no major predominance of any individual specific regulatory mechanism observed (Additional file 6: Table S3).

A bionetwork cluster analysis was performed on the differentially expressed genes in the various brain regions as previously described [29,31] to identify gene modules and networks with coordinated and interconnected relationships (i.e. connectivity) [38] (Figure 1). Initially all the differentially expressed genes in the combined brain regions for male or female were analyzed, termed combined networks (Figure 3a). This combined analysis was performed to potentially identify common gene networks or modules similar among all the brain regions that potentially correlate with the mate preference behavior parameters. The increased number of microarrays and data associated with the combined analysis also improves the power of the cluster and network analysis. The gene cluster analysis is shown and individual modules of genes identified are presented in different colors on the axis. The module colors represent increasing levels of connectivity [38] with white being negligible and red being highest. The combined Signature lists provided 4 modules in the female and 9 gene modules in the male. This can be seen as the blocked gene clusters designated as modules of different colors (Figure 3a). The number of genes in each module for male and female brain regions is shown in Table 1. This combined network analysis and modules were correlated with the mate preference behavior, but no significant correlations were found using this combined analysis (Additional file 7: Table S4).

Due to the distinct functions of each brain region and distinct gene Signature lists, the combined analysis was found not to provide the specificity needed to identify the behavioral correlations with gene modules. Therefore, a more specific network analysis using the individual Signature lists for each brain region separately was performed, termed separate networks. Each brain region differentially expressed gene Signature list was used for separate network analysis. The cluster analysis identified specific gene modules for each brain region from the male and female gene Signature lists presented (Figure 3b, c). The modularity for the specific brain regions was not as strong as the combined region analysis. Each separate brain region is shown and the gene modules are identified by the different colors. The brain regions had 1–6 different modules

Brain Regions	Genes Number in Signature List	Number of KEGG Pathways	6 Female Regions	F-Amy	F-CngCTX	F-EnCTX	F-Hipp	F-OlfB	F-POAH	6 Male Regions	M-Amy	M-CngCTX	M-EnCTX	M-Hipp	M-OlfB	M-POAH	Signature Lists Genes Overlap
6 Female Regions	**1833**	*161*		139	803	433	70	748	56	230	36	93	46	24	50	13	
F-Amy	139	*18*	*18*		42	6	1	20	2	23	5	9	6	2	4	2	
F-CngCTX	803	*108*	*108*	*18*		45	1	245	7	63	2	33	15	8	7	4	
F-EnCTX	433	*128*	*126*	*12*	*90*		9	31	11	83	19	34	16	7	21	3	
F-Hipp	70	*34*	*34*	*9*	*20*	*28*		9	3	11	2	3	0	1	6	1	
F-OlfB	748	*123*	*123*	*16*	*89*	*98*	*26*		9	82	12	29	13	9	21	2	
F-POAH	56	*47*	*47*	*9*	*35*	*42*	*14*	*44*		13	3	2	3	3	3	5	
6 Male Regions	**1693**	*161*	*161*	*18*	*108*	*126*	*34*	*123*	*47*		175	785	385	151	356	43	
M-Amy	175	*55*	*55*	*12*	*43*	*47*	*18*	*48*	*30*	*55*		27	11	11	19	1	
M-CngCTX	785	*153*	*153*	*17*	*106*	*122*	*32*	*119*	*47*	*153*	*53*		36	1	51	6	
M-EnCTX	385	*150*	*150*	*18*	*104*	*123*	*33*	*115*	*47*	*150*	*52*	*147*		6	12	4	
M-Hipp	151	*121*	*121*	*16*	*86*	*100*	*28*	*95*	*45*	*121*	*47*	*118*	*119*		22	3	
M-OlfB	356	*142*	*142*	*14*	*93*	*114*	*34*	*109*	*45*	*142*	*51*	*135*	*134*	*119*		5	
M-POAH	43	*13*	*13*	*2*	*9*	*11*	*2*	*7*	*5*	*13*	*4*	*13*	*13*	*11*	*13*		
			Enriched KEGG Pathways Overlap														

Figure 2 Number of overlapped differentially expressed genes with pathways and Signature lists. Number of genes overlapped between Signature lists is shown in regular font on grey background; number of affected KEGG pathways overlapped between Signature lists is shown in italicized font in white background; only KEGG pathways with 5 or more genes affected are counted.

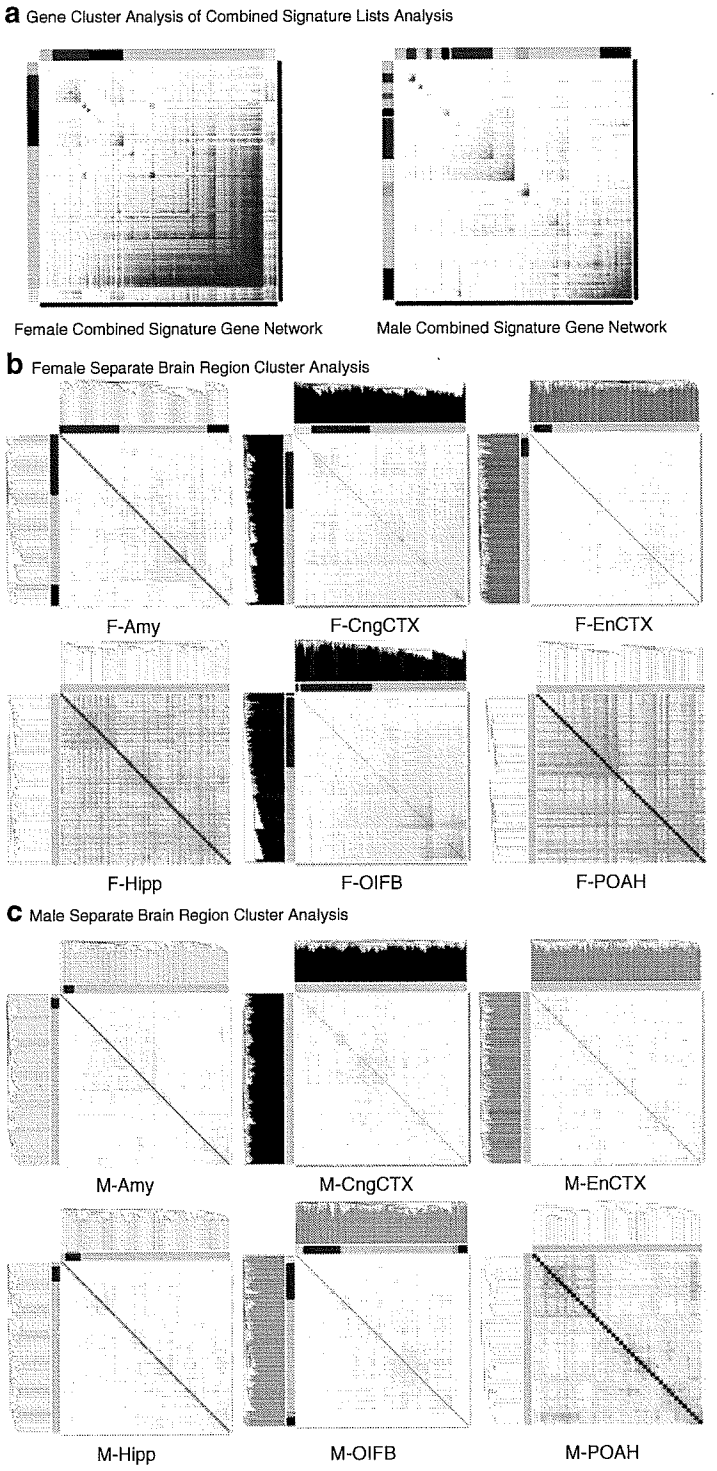

Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 Male and female brain region bionetwork cluster analysis and corresponding gene modules. Topological overlap matrixes of the gene co-expression network consisting of genes differentially expressed in F3 generation vinclozolin lineage rat brain as compared to F3 generation lineage Control animals. Genes in the rows and columns are sorted by an agglomerative hierarchical clustering algorithm. The different shades of color signify the strength of the connections between the nodes (from white signifying not significantly correlated to red signifying highly significantly correlated). Modules identified are colored along both column and row and are boxed. **(a)** Matrixes of combined network for male and female brain regions. **(b)** Matrixes of separate network for female amygdala (F-Amy), cingulate cortex (F-CngCTX), enterorhinal cortex (F-EnCTX), hippocampus (F-Hipp), olfactory bulbs (F-OlfB), and preoptic area-anterior hypothalamus (F-POAH). **(c)** Matrixes of separate network for male amygdala (M-Amy), cingulate cortex (M-CngCTX), enterorhinal cortex (M-EnCTX), hippocampus (M-Hipp), olfactory bulbs (M-OlfB), and preoptic area-anterior hypothalamus (M-POAH).

and associated gene networks (Table 1). The same differentially expressed gene Signature lists were used, but the network analysis was from the separate lists (Figure 3b and c). All subsequent analyses used the gene modules from this region specific network analyses.

The cluster analysis (Figure 3b and c) for each brain region provided modules of genes with coordinated gene expression and identified a connectivity index [29-31,33,38] for each associated gene. The connectivity index (k.in) for each of the differentially expressed genes in each region is presented in Additional file 4: Table S2. The top 10% of genes with the highest connectivity index for each of the gene modules was identified (Additional file 4: Table S2 as the genes in bold font). From this combined list of 185 genes for male and 225 genes for female, a gene sub-network analysis was performed. The most highly interconnected genes in all modules for female and male brain regions were used to identify the common direct connection interactions between genes in a gene sub-network (Figure 4). The female gene sub-network identified angiogenesis, growth and apoptosis as predominant cellular processes affected (Figure 4b). The male gene sub-network identified apoptosis as a predominant pathway affected (Figure 4a). These gene networks identify the common connections within the brain regions with the most highly interconnected genes differentially expressed between the control and vinclozolin F3 generation animals.

Analysis of the gene networks for each individual brain region gene module demonstrated that only males exhibited direct connection gene sub-networks for gene modules (male amygdala and cingulate cortex turquoise modules) (Figure 5). This region specific examination of gene network modules demonstrated most regions did not have direct connection sub-networks, but indirect interactions with various pathways and processes. An alternative analysis used the entire Signature list for each brain region to identify region specific gene sub-networks (Additional file 8: Figure S2 A-K). How these gene networks may correlate with the alterations in mate preference behavior required a statistical correlation of the gene sets with the behaviors (Figure 1).

In considering the mate preference behaviors, the female is the discriminatory sex to choose a mate, while the male is non-discriminatory and has phenotypes and behaviors to be selected [27,39]. The altered gene expression and correlations with behaviors needs to consider this in data interpretation. The behavioral parameters (Additional file 2: Table S1) for the mate preference analysis were statistically correlated to the separate network gene modules for the different brain regions, (Additional file 9: Table S5). The correlation and the p-values associated with the statistical correlation coefficients are presented. All correlations with a single or multiple principle component comparison are presented. Considering a $p < 0.05$ or correlation coefficient >0.5 and $p = 0.05$-0.1 between the gene module and behavior demonstrated correlations in four female brain regions and six modules with the female behavior (Table 2 and Additional file 9: Table S5). A summary of the statistically significant correlations and/or those with strong correlation coefficients is shown in Figure 6. Nearly all the female brain regions had statistically significant correlation with the Plexiglas behavioral parameter. The female amygdala (F-Amy) had a turquoise module with significant correlation with the Walking and Still parameters. The turquoise modules of female entorhinal cortex (F-EnCtx) had a strong correlation with the Wire Mesh and Plexiglas behavioral parameters (Figure 6).

The six male brain regions and associated gene modules had a number of statistically significant correlations with the mate preference parameters (Table 2 and Additional file 9: Table S5). All the male brain regions had statistically significant correlation with at least one module and the Wire Mesh behavioral parameter. Amy and Hipp also had correlations with the behavioral Plexiglas parameter (Figure 6). Therefore, at least one gene module in nearly all brain regions statistically correlated to the mate preference parameters analyzed. These correlations can now be considered in regards to the regulatory roles of gene networks identified for mate preference behavior alterations for the female (chooser) versus the male (selected) (Figure 5 and Additional file 8: Figure S2).

The direct connection gene sub-networks for the critical male amygdala (M-Amy) and cingulate cortex (M-CngCtx) turquoise modules are shown in Figure 5. The Signature list for each brain region sub-networks demonstrate distinct networks for each region (Additional file 8: Figure S2). Since nearly all the brain regions and key modules (Figure 6) have a statistically significant correlation with the Wire

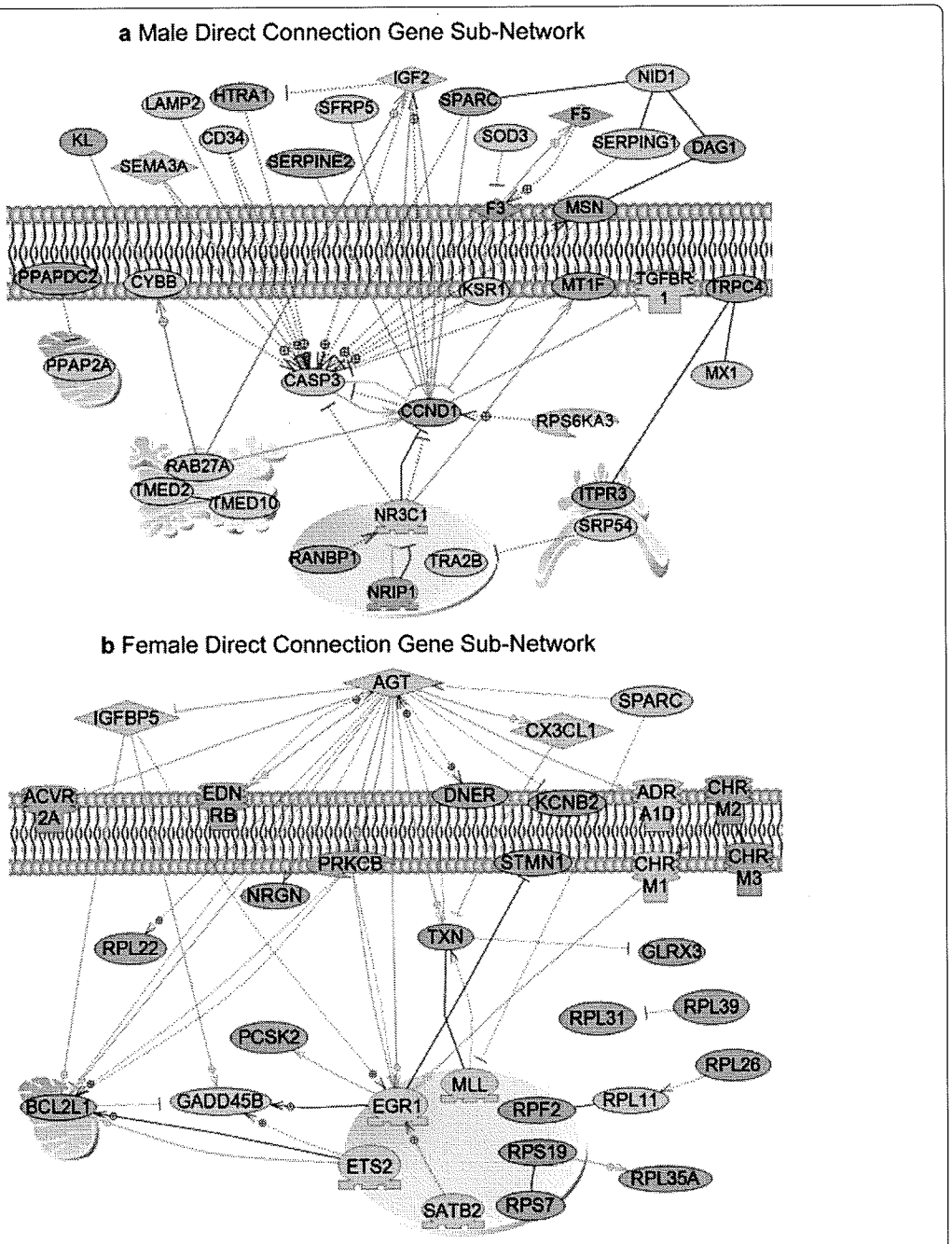

Figure 4 Direct connection sub-networks for most highly top 10% connected genes from each module of separate network for male (a) or female (b). Only directly connected genes are shown according to their location in the cell (on membrane, in Golgi apparatus, nucleus, cytoplasm or outside the cell). Node shapes and color code: oval and circle – protein; diamond – ligand; circle/oval on tripod platform – transcription factor; ice cream cone – receptor; crescent – kinase or protein kinase; irregular polygon – phosphatase; red color indicates up-regulated genes, blue – down-regulated. Arrows with plus sign show positive regulation/activation, arrows with minus sign – negative regulation/inhibition; grey arrows represent regulation, lilac - expression, purple – binding, green – promoter binding, and yellow – protein modification.

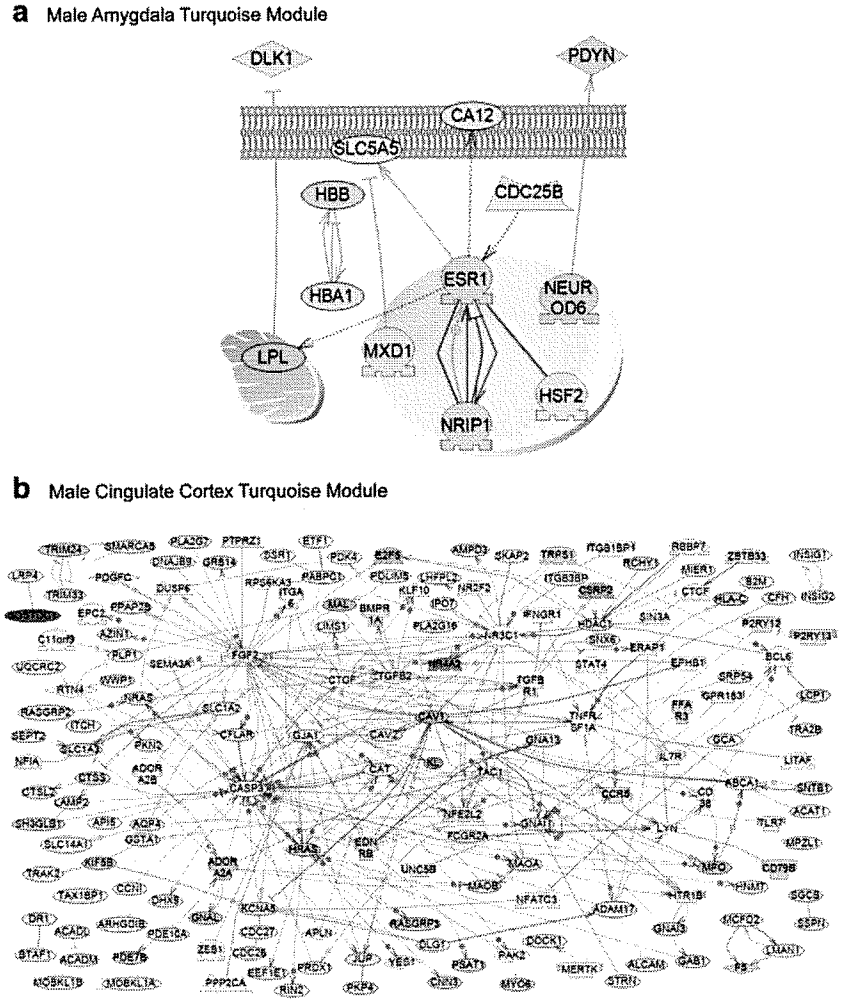

Figure 5 Direct connection sub-networks for male amygdala turquoise module (a) and male cingulate cortex turquoise module (b). Only directly connected genes are shown. Node and arrows shapes and color code is the same as for Figure 4.

Mesh for male or Plexiglas for female mate preference behavior parameters, the combined gene sub-network (Figure 4) for all male or female brain regions identifies a potentially associated molecular control of behavior. Alternately, the analysis of separate regions differentially expressed gene sets (Signature lists) identified distinct gene sub-networks that associate with the different regions (Additional file 8: Figure S2). These potential gene sub-networks correlate and potentially regulate the mate choice behavior for the female and selection behavior/phenotype for the male. In addition to the gene networks, a correlation of critical cellular pathways in specific brain regions and modules (Additional file 5) that are associated with the epigenetic transgenerational inheritance of altered mate preference behavior.

The epigenetic transgenerational inheritance of the altered mate preference behavior requires the transmission of an altered epigenome in the germline (sperm) [5,13,15]. Previously the altered DNA methylation of the F3 generation sperm was characterized with 48 differentially DNA methylated regions (DMR) being identified in gene promoters [14]. These vinclozolin induced sperm DMR are in part what promotes an altered epigenome in the embryo and all developing tissues transgenerationally [15]. Although any developing tissue (e.g. brain) will have a dramatic cascade of epigenetic and genetic steps to achieve an adult fully differentiated state [40,41], the possibility that some of the original germline epigenetic marks (DMR) may persist was investigated. The genes associated with the 48 previously identified sperm DMR

Table 2 Gene modules highly correlated to different mate preference behavior parameters

Sex-region	Behavior trait		Wire mesh		Facial		Plexiglas		Still		Walking	
	Module	# PC*	Correlation	p-value	Correlation	p-value	Correlation	p-value	Correlation	p-value	Correlation	p-value
F-Amy	Turquoise	2					0.52	**0.028**	0.82	**0.012**	0.90	**0.001**
F-EnCTX	Blue	1					0.55	**0.066**				
	Brown	1					0.41	**0.037**				
	Turquoise	3	0.81	**0.029**			−0.53	**0.083**				
F-OlfB	Yellow	2					0.74	**0.044**				
F-POAH	Turquoise	1					0.63	**0.0386**				
M-Amy	Blue	2					0.81	**0.008**				
	Turquoise	1	−0.51	**0.022**								
M-CngCTX	Turquoise	1	0.54	**0.048**								
M-EnCTX	Turquoise	1	−0.58	**0.037**								
M-Hipp	Blue	1	0.57	**0.023**			0.77	**0.017**				
	Turquoise	1	0.60	**0.034**								
M-OlfB	Blue	2	0.79	**0.012**							0.67	**0.072**
	Brown	1	0.73	**0.012**								
	Green	1	0.62	**0.020**								
	Red	1			0.55	**0.081**						
	Yellow	1	−0.60	**0.033**								
M-POAH	Turquoise	1	0.87	**0.002**							0.753	**0.091**

*- number of principal components (PC) used to calculate correlation between modules and behavior.

were compared with the male and female brain region gene Signature lists. The comparison demonstrated the majority of the DMR did not correspond to differentially expressed genes in various brain regions. Only Rnase1 in the male Amy, Ig6-2a in the male EnCtx, Parp9 in the female CngCtx and Rp132 in the female OlfB overlapped. Interestingly, a copy number variation (CNV) in the Fam111a site previously identified [14] was found in all brain regions identified with the epigenome analysis. This provides a positive control for the technology and ability to detect the DMR. Therefore, some of the original sperm DMR programmed sites may

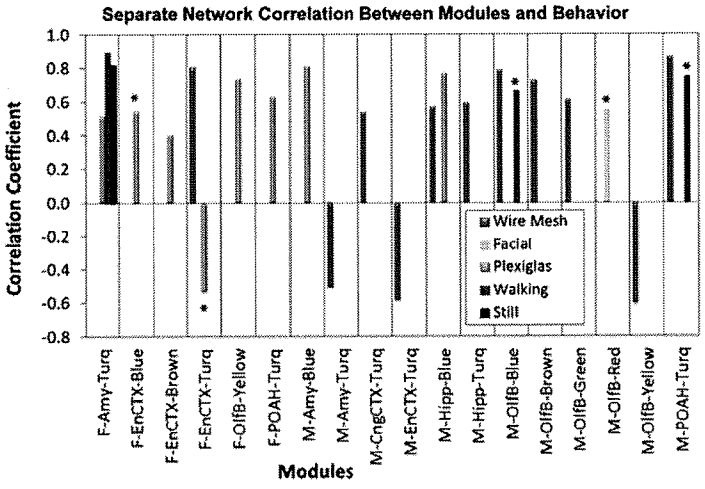

Figure 6 Gene module correlation with mate preference behavior parameters. Separate network male and female modules highly correlated to types of behavior: Wire Mesh (blue bars), Facial Investigation (orange), Plexiglas (green), Walking (red) and Still (black). Bars not marked with asterisks have p-value < 0.05; bars marked with one asterisk have correlation coefficient >0.5 and > p-value = 0.1 - 0.05.

persist, but the vast majority of brain development and epigenetic programming, and potential distal regularity role of DMR in epigenetic control regions [25], is distinct from the original germline epigenetic marks.

Discussion

A systems biology analysis of environmentally induced epigenetic transgenerational inheritance of altered mate preference behavior was performed to suggest a potential role for epigenetics in evolutionary biology. Previous research has demonstrated that environmental toxicants such as the fungicide vinclozolin can promote a reprogramming of the germline epigenome during fetal gonadal sex determination that then transmits altered phenotypes and adult onset disease states transgenerationally in the absence of future environmental exposure [41]. This is referred to as epigenetic transgenerational inheritance [5,15] and suggests a role for environmental epigenetics in the inheritance of phenotypic variation and disease, independent of classic genetic inheritance mechanisms. The basic molecular mechanism involved in this non-genetic form of inheritance is the ability of environmental factors to influence the epigenetic programming of the germline [15,19]. The primordial germ cells during migration down the genital ridge undergo an erasure of DNA methylation that then is initiated to re-methylate at the time of gonadal sex determination in a sex-specific manner [19]. An environmental toxicant such as vinclozolin appears to alter gonadal development to influence germline DNA methylation programming [42] and the differential DNA methylation regions (DMR) in the sperm become imprinted-like sites that appear to not get erased at fertilization so are transmitted to subsequent generations and male and female progeny [13-16]. In addition to vinclozolin, a number of other environmental toxicants such as the plastic compound bisphenol A (BPA) [16,43], dioxin [16,44], methoxycholor [13], phthalates [16], pesticides [16], hydrocarbons [16], and DDT [17] have been shown to induce transgenerational phenotypes. Other environmental factors such as nutrition and stress can also promote transgenerational phenotypes [45-48].

The vinclozolin induced epigenetic transgenerational phenotypes previously identified included adult onset rat disease after 12 months of age of male infertility, mammary tumors, prostate disease, kidney disease and immune abnormalities [20]. Therefore, the mate preference analysis was performed prior to adult onset disease to remove the disease as a confounding factor. A brain-behavior transgenerational phenotype observed was increased female anxiety and decreased male anxiety behaviors [23]. This transgenerational anxiety behavior was also examined on a molecular level to identify brain region specific changes in different gene expression and gene networks associated with the behavior [23]. Similar observations were made in the analysis of transgenerational stress responses [26]. Interestingly, in a previous study we found that vinclozolin induced alterations in mate preference behavior [27]. Females from either control or vinclozolin F3 generation lineages prefer control lineage males over vinclozolin lineage males, whereas no altered mate preference in males was observed [27]. While the standard argument would be that the females are the discriminating sex and distinguish between males on as yet undetermined phenotype characteristic(s), it is important to realize that the absence of evidence (in the male) does not mean the evidence of absence of male involvement since preference is only the first step in a mating sequence. That is, under natural circumstances this is followed by a mutual decision. Mating in rodents involves pheromone and auditory cues produced by both sexes and evidence suggests (see below) that it is under such unfettered conditions that the complementarity of behavior and brain are expressed.

Observations from the current study need to consider the effects on the female brain as potentially altering female discrimination and preference. The effects on the male brain are presumed to be associated with the characteristics (e.g. auditory cues and pheromone production) being selected. This altered mate preference behavior suggests the existence of an environmentally altered epigenetic transgenerational inheritance of mate preference behavior [15]. The current study was designed to identify the gene bionetworks in various male and female brain regions that correlate with the behavior of the transgenerational inheritance model.

A novel gene bionetwork analysis was developed to identify gene networks correlated to disease [29]. The approach was to use a large number of microarrays to identify transcriptomes in specific tissues associated with control versus disease individuals in large cohorts. Differentially regulated genes that are coordinately regulated and having connectivity [38] are clustered in large gene sets to identify modules of genes that associate with the disease [29,30,33-35] (Figure 1). More recently, we have used a similar approach to investigate a normal development process to identify gene bionetworks associated with development [31,32]. The primordial follicle development in the ovary was investigated to identify a network of growth factors and associated signaling systems that regulate follicle development [31,32]. This bioinformatics approach to identify regulatory gene networks was used in the current study to correlate brain gene networks to mate preference behavior (Figure 1) in an epigenetic transgenerational model [27]. The six different brain regions isolated from F3 generation control and vinclozolin lineage females and males were used in a microarray analysis to determine the differential gene expression in each brain region. The region specific gene sets, "Signature list", and associated gene networks were investigated.

Analysis of significant pathways and cellular processes potentially influenced by the differentially expressed gene sets and networks did not identify predominant or over represented pathways. One pathway identified that previously has been shown to correlate with sexual selection is the olfactory transduction pathway [49-52]. However, most of the brain regions and specific networks or gene modules affected similar pathways with overlap between most. Therefore, no specific pathways were identified and most major pathways were influenced. A limitation in gene expression studies is that individual genes are assigned a specific function, but the gene may be involved in functional categories. This limitation needs to be considered in any gene expression data interpretation, but genome wide transcriptome analysis for gene sets has been shown to start to address this issue [53]. Combined observations suggest that the distinct differential expressed gene sets for the different brain regions appear to regulate common cellular processes and pathways among the brain regions and modules.

In contrast, analysis of gene networks identified unique gene sub-networks and gene modules associated with each brain region differentially expressed gene sets. The specific inter-connected genes were unique and overall networks of connected genes distinct. Therefore, the different functions associated with each brain region and associated with the altered mate preference behavior were identified. A statistical correlation of the gene modules for each brain region with the different mate preference parameters measured identified a number of statistically significant correlations. All but one female brain region (Hipp) had statistically significant correlations with the Plexiglas behavior parameter. The female behavior directly associates with the discrimination and mate preference choice. Interestingly, all the male brain regions had some modules with statistically significant correlations with the Wire Mesh behavior parameter. The male behavior and/or phenotype (e.g. pheromonal production) associates with the selected behaviors and characteristics of the non-discriminant sex. Therefore, direct correlations with the gene modules, specific brain regions and mate preference behavior parameters were identified. The specific gene modules and behavioral parameters statistically correlated were distinct between the sexes and brain regions, but strong correlations of the gene networks to the mate preference behavior was established. Interestingly, both the Plexiglas and Wire Mesh are indicators of interest and assessment of the stimulus animal.

The gene bionetwork analysis and statistical correlation with the mate preference behavior provides insight into the molecular basis of how various male and female brain regions correlate and in part control the various behavioral parameters. Observations provide one of the first genomic and systems biology analysis of mate preference behavior (Figures 1 and 7). The experimental model used

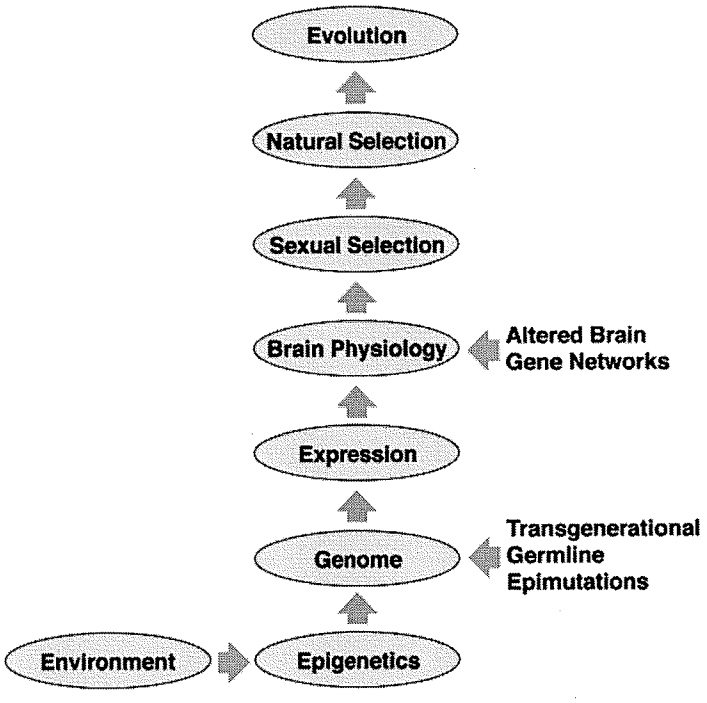

Figure 7 Schematic of role of epigenetics in evolution.

involved the ability of an environmental compound (vinclozolin) to induce an epigenetic modification of the germline (sperm) to promote epigenetic transgenerational inheritance of an altered mate preference behavior. This altered mate preference behavior was due to a baseline alteration in the epigenomes of all male and female tissues, including the brain, which are derived from the epigenetically altered germline [25]. The current study used a systems biology approach to help elucidate the molecular control of this process.

Darwin [12] considered natural selection and sexual selection as distinct processes driving the evolution of traits. Natural selection results in traits that are adaptive responses to changes in the environment. The resulting variation in traits between and within species is shaped by differential survivorship. In other words, animals that survive are those with traits that are adaptive to their environment [54]. Darwin [12] conceived of sexual selection as arising from aggressive interactions between males (male-male competition) and the female's selection of a mate (mate choice). Males compete amongst themselves for access to females. Aggression between males can have a direct effect on female reproduction by preventing other breeding males from having access to females or from harming the female. Importantly aggression amongst males can also have an indirect effect by inhibiting or suppressing the normal reproductive physiology of the female or even terminate a pregnancy [55].

Mate preference, in its simplest form, states that males compete for females and females choose between them. Although most research has focused on how females choose males, male choice of females is also important [56-58]. This point cannot be overemphasized. That is, in virtually all paradigms published to date, the choosing individual is the independent variable and the stimulus animal is the dependent variable. Although this study is similar to previous studies in that there are restrained stimulus animals and freely moving individuals that are the investigators, it differs in several important ways. First, the 'round robin' testing method insured that all males and all females served both as stimulus animals and experimental animals; thus, the "preferences" exhibited reflect both the males and the females. Second, this study extends to the molecular level events in the brain of the individuals, both of which have exhibited preferences, namely brain transgenerational transcriptome alterations that correlated to the opposing sex behaviors. In essence we are seeing the outcome of the complementary nature of mate choice. This has never been shown before.

Evolution favors reproductive success, and it is in the individual's interest to focus on selecting the best mate and to avoid mating with the wrong species [59,60]. Making the correct choice of a mate has a pronounced impact on reproductive success of both partners. Except in unusual systems, in nature the mating partners choose one another [55,57,60]. Experiments with flies [61], birds [62], and rodents [63,64] indicate that individuals who are allowed to select, and be selected by, their mate enjoy greater reproductive success than force-paired animals. This consent is based not only on the internal milieu that motivates each individual to seek a partner, but also on the satisfactory nature of the phenotypic traits the potential mate displays.

There are a number of sexual selection hypotheses, all of which emphasize that females choosing optimally will produce young whose viability and survivorship are enhanced by the female's choice of mate [65]. The most attractive, and one that takes into account that mating is a cooperative act that involves both partners, is the sensory exploitation hypothesis [66]. This hypothesis postulates that males have evolved calls and/or pheromones to exploit the preexisting sensory biases in the female that themselves evolved for reasons independent of female choice. Male behavior then has changed to maximize stimulation of the female's sensory systems. Therefore, the coordination of the complementary signal and receiver, mounting and lordosis, coordination of egg and sperm maturation and release is required for successful completion for reproduction. These complementary processes are evident at all levels of biological organization [55,67] and we extend it here to the level of the genome and epigenome.

Conclusions

The ability of an environmental factor to alter mate preference behavior suggests a critical role of environment in evolutionary biology. This is distinct from the generally accepted role of the environment in natural selection where environment is the active factor in the selection of an adaptive phenotype, but alternately here it involves the induction of phenotypes that can be acted on by natural selection. Since the majority of environmental factors can not alter DNA sequence or promote mutagenesis [2], an additional molecular mechanism to consider involves environmental epigenetics [15]. Many environmental compounds and factors such as nutrition can modify the epigenome to alter phenotypic variation. The role of epigenetics in evolutionary biology has been suggested previously [5,8-11,15,68], but no significant experimental evidence has been provided. The current study demonstrates an environmental factor can promote the epigenetic transgenerational inheritance of an altered mate preference behavior. The epigenetic modification of the germline (sperm) has been previously established [13,14,16] and will lead to epigenetic alterations in the brain transcriptomes of both females and males [23] to alter the mate preference behavior [27]. Therefore, the current study provides direct experimental evidence for a potential role of environmental

epigenetics in evolution by regulating a critical determinant such as mate preference on a molecular level (i.e. altered gene networks) in specific brain regions in a sex-specific manner (Figure 7). Although no direct epigenetic alterations were examined in the brain, the germline (sperm) epigenetic alterations that generate this altered male or female brain development have been documented [13,14,16]. This molecular mechanism does not suggest genetics will not have a critical role in evolutionary biology, but suggests environmental epigenetics will be an additional mechanism to consider. Epigenetics provides a mechanism for the environment to impact phenotypic variation and natural selection. Epigenetic and genetic mechanisms will cooperate to regulate on a molecular basis evolutionary biology. This appears to be a "neo-Lamarckian concept to facilitate neo-Darwinian evolution" [40,41].

The systems biology approach used in the current study links an environmental exposure, epigenetic transgenerational inheritance and molecular regulation of brain function to mate preference and evolutionary biology (Figure 7). Epigenetics will have a central role in how environmental factors influence how the gene networks emerge to induce phenotypic variation. Although genetics is critical for all aspects of biology, epigenetics provides the plasticity to allow the environment to alter biological events. This type of systems approach to understand complex biological traits, such as sexual selection, provides insights into how the various components (environment, phenotype and evolution) interact in a systems biology manner.

Methods

Animal housing protocol

Male and female rats of the F3 generation of Vinclozolin (Vinclozolin-Lineage) and DMSO Control (Control-Lineage) Lineages were selected out of litters from untreated F2 generation mothers in Dr. Michael Skinner's laboratory at Washington State University according to established protocols [13]. Briefly, gestating female F0 generation Sprague Dawley rats were injected with the fungicide vinclozolin (100 mg/kg) daily during fetal gonadal sex determination (E8-E14) and the F1 generation were bred to generate the F2 generation and then the F2 bred to generate the F3 generation [13]. At approximately PND 10 (before weaning), each animal was injected with a small microchip (AVID Identification system Inc. Norco, CA) subcutaneously between the shoulder blades. The animals were then shipped to the University of Texas from Washington State University on postnatal day (PND) 22, one day after weaning. Upon arriving at the University of Texas, one animal from each Lineage (Control and Vinclozolin) was pair-housed (one control and one vinclozolin animal) and remained in these dyads throughout the duration of the study. Because of the natural variation in dates of breeding, there was a 4-day spread of birth date of animals in the first cohort but in the second cohort, all animals were born on the same day. However, all pair-housed animals were no more than one day apart in birth age and were paired randomly to prevent an age effect on cagemates.

Each dyad of animals was randomly placed in a six-wide, five-high metal housing rack in standard translucent polycarbonate rat cages (46 × 24 × 20.5 cm) with *ad libitum* access to tap water and standard rat chow (Purina rodent chow #5LL2 Prolab RMH 1800 diet). The animal room was on a 14:10 light/dark schedule. For environmental enrichment, a 7 cm diameter PVC pipe was placed in each cage.

Ethics statement

All experimental protocols for the procedures with rats were pre-approved by the Washington State University Animal Care and Use Committee (IACUC approval # 02568–026) and by the University of Texas at Austin Animal Care and Use Committee (Public Health Service Animal Welfare Assurance Number A4107-01).

Behavioral testing

After habituation to the testing arena, each individual was tested individually (when used as an experimental subject) or in pairs (when used as stimulus animals) with all individuals; the order of the testing was rotated during the course of both the male and female trials. All tests were conducted during the dark phase of the light cycle, beginning at 1200 h, 4 h after the progesterone injection, in a room illuminated with low levels of red light. Before trials, to confirm that females were receptive, each female was placed with a sexually experienced but otherwise experimentally naïve male; all females exhibited robust lordosis (arched back and lifted head posture) in response to mounting by the male.

Partner preference tests consisted of placing an individual (male or female) in the center of a large three-chamber glass-testing arena (122 × 46 × 54 cm). At either end was a small compartment (28 × 28 × 12.5 cm) containing the stimulus rats separated by a Wire-mesh barrier to allow exchange of olfactory, visual, and tactile cues. The area directly in front of the stimulus cage was marked by tape. Tests were conducted 2 h after the onset of the dark cycle under red-light illumination and lasted 10 min; all tests were videotaped for further review and analysis. At the end of each test, all animals were removed, and the entire testing arena was washed with a household cleaner and then wiped down with 70% ethanol to remove scent marks and residual odors. All males were tested with both types of females as stimulus animals (72 trials), and all females were tested with both types of males as stimulus animals (72 trials) (Movie S1).

The videotaped trials were analyzed by using JWatcher v1.0 (www.jwatcher.ucla.edu) computer software to quantify the behavior of each experimental animal. Time spent with a stimulus animal was recorded as soon as all four paws of the experimental animal crossed over the line of tape marking the boundary of that stimulus animal's compartment. As soon as one paw crossed over the tape back into the center compartment, the time recorded with the experimental animal was stopped. Preference behaviors were defined as those directed to the stimulus animal and included time spent in contact with the Wire Mesh separating the experimental and stimulus animal (Wire Mesh), during which the animals often touched noses through the Wire Mesh (facial investigation), and contacted the Plexiglas surface surrounding the front of the stimulus cage; the cumulative total time in these preference behaviors toward each stimulus animal was also calculated (Total). Other activity measured included undirected walking and sniffing (walking), standing still with minimal head movement (still). Videos demonstrating the test can be viewed as supporting information (Movie S1).

Brain processing

The brain was removed in less than 1 minute and placed in crushed ice to chill. The brain was then cut in half in the sagittal plane along the midline. In all cases but one the right side was blocked and then 6 areas dissected (see list below) within 3–5 min. This procedure was done on iced tissue. The dissected brain areas were placed in chilled Trizol (150 l) in 1.5 ml Eppendorf tubes according to manufacturers specifications in each tube. No tissue fragment was more than 3 mm but in those instances, but where there were multiple fragments the amount of Trizol was doubled (approximately). After all animals were dissected, the Eppendorf tubes were vortexed (15 sec) and then frozen on dry ice. The brain regions collected were according to Paxinos & Watson [69]: olfactory bulbs (OlfB); cingulate cortex (CngCtx), anterior to POAH (Bregma 4.7 to 1.7); preoptic area-anterior hypothalamus (POAH), 4 mm rostral to anterior commissure (AC) (Bregma −0.26 to −1.40); amygdaloid nuclei (Amy), 3 mm caudal to AC (Bregma −2.3 to −3.6); hippocampus (Hipp), 6 mm caudal to AC (Bregma −2.12 to −4.52); entorhinal cortex (EnCtx) (Bregma −5.60 to −7.80).

RNA preparation

Brain area samples from individual rats were homogenized in 150 μl Trizol and then 600 μl Trizol was added to final volume of 750 ml. Samples were stored at −80 or −20°C until RNA extraction. For microarray analysis, from 4 to 6 biological replicas (animals) were prepared as above for each brain area Control or Vinclozolin group depending on samples availability (Additional file 2: Table S1B). A total of 132 (67 Control and 65 Vinclozolin) samples/chips were analyzed: (6 brain areas) × (2 Male or Female) × (2 Control or Vinclozolin) × (4–6 biological replicas). RNA from individual animal brain area was extracted from Trizol samples according to standard Trizol extraction protocol (Invitrogen, USA) and stored in aqueous solution at −80°C until microarray analysis.

Microarray analysis

The microarray analysis was performed by the Genomics Core Laboratory, Center for Reproductive Biology, Washington State University, Pullman, WA using standard Affymetrix reagents and protocol. Briefly, mRNA was transcribed into cDNA with random primers, cRNA was transcribed, and single-stranded sense DNA was synthesized which was fragmented and labeled with biotin. Biotin-labeled ssDNA was then hybridized to the Rat Gene 1.0 ST microarrays containing more than 30,000 transcripts (Affymetrix, Santa Clara, CA, USA). Hybridized chips were scanned on Affymetrix Scanner 3000. CEL files containing raw data were then pre-processed and analyzed with Partek Genomic Suite 6.5 software (Partek Incorporated, St. Louis, MO) using an RMA, GC-content adjusted algorithm. Raw data pre-processing was performed in 12 groups, one for each male or female brain area. Comparison of array sample histogram graphs for each group showed if data for all chips were similar and appropriate for further analysis (Additional file 3: Figure S1). By this criterion, 2 microarray samples (not counted in Additional file 4: Table S2B and not shown on Additional file 3: Figure S1) were omitted from repeated group pre-processing and further analysis.

The microarray quantitative data involves signals from an average of 28 different oligonucleotides (probes) arrayed for each transcript and many genes are represented on the chip by several transcripts. The hybridization to each probe must be consistent to allow a statistically significant quantitative measure of resulting gene expression signal. Therefore, the microarray provides an unbiased and highly stringent quantitative procedure compared to other protocols [70]. In contrast, a quantitative PCR procedure uses only two oligonucleotides and primer bias is a major factor in this type of analysis. Therefore, we did not attempt to use PCR based approaches as we feel the microarray analysis is more accurate and reproducible without primer bias such as PCR based approaches [31].

All microarray CEL files from this study have been deposited with the NCBI gene expression and hybridization array data repository GEO (GEO series accession number: GSE33830) and can be also accessed through www.skinner.wsu.edu. For gene annotation, Affymetrix annotation file RaGene1_0stv1.na31.rn4.transcript.csv was used.

Network analysis

The network analysis was restricted to genes differentially expressed between the control and the treatment

groups based on previously established criteria: (1) fold change of group means ≥ 1.2 or ≤ 0.83; (2) *T* test p-value ≤ 0.05. The union of the differentially expressed genes from the different treatments resulted in 1,693 genes for males and 1833 for females being identified and used for constructing a weighted gene co-expression network [71,72]. Unlike traditional un-weighted gene co-expression networks in which two genes (nodes) are either connected or disconnected, the weighted gene co-expression network analysis assigns a connection weight to each gene pair using soft-thresholding and thus is robust to parameter selection. The weighted network analysis begins with a matrix of the Pearson correlations between all gene pairs, then converts the correlation matrix into an adjacency matrix using a power function $f(x) = x^{\beta}$. The parameter β of the power function is determined in such a way that the resulting adjacency matrix (i.e., the weighted co-expression network) is approximately scale-free. To measure how well a network satisfies a scale-free topology, we use the fitting index proposed by Zhang & Horvath [71] (i.e., the model fitting index R^2 of the linear model that regresses *log(p(k))* on *log(k)* where k is connectivity and *p(k)* is the frequency distribution of connectivity). The fitting index of a perfect scale-free network is 1.

To explore the modular structures of the co-expression network, the adjacency matrix is further transformed into a topological overlap matrix [73]. As the topological overlap between two genes reflects not only their direct interaction, but also their indirect interactions through all the other genes in the network. Previous studies [71,73] have shown that topological overlap leads to more cohesive and biologically meaningful modules. To identify modules of highly co-regulated genes, we used average linkage hierarchical clustering to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules [74]. Such networks were generated from all combined male or female differentially expressed genes (2 combined networks) or from each individual male or female brain region Signature lists (12 separate networks). From one to ten modules were identified in combined or separate networks and the module size was observed to range from 10 to 780 genes (Table 1).

To distinguish between modules, each module was assigned a unique color identifier, with the remaining, poorly connected genes colored grey. The hierarchical clustering over the topological overlap matrix (TOM) and the identified modules is shown (Figure 1). In this type of map, the rows and the columns represent genes in a symmetric fashion, and the color intensity represents the interaction strength between genes. This TOM heatmap highlights that genes in the transcriptional network fall into distinct network modules, where genes within a given module are more interconnected with each other (blocks along the diagonal of the matrix) than with genes in other modules. Therefore, there are two types of global connectivity, adjacency-based one and TO based one. The adjacency-based connectivity (*k.all*) is defined as the sum of the power-function transformed correlations between the gene *g* and all the other genes in the whole network while the TO-based connectivity (*to.all*) is defined as the sum of the topological overlaps between the gene *g* and all the other genes. By default, connectivity used throughout the paper refers to TO-based connectivity *to.all*.

Gene Co-expression Network Analysis Clarification: Gene networks provide a convenient framework for exploring the context within which single genes operate. Networks are simply graphical models comprised of nodes and edges. For gene co-expression networks, an edge between two genes may indicate that the corresponding expression traits are correlated in a given population of interest. Depending on whether the interaction strength of two genes is considered, there are two different approaches for analyzing gene co-expression networks: 1) an unweighted network analysis that involves setting hard thresholds on the significance of the interactions, and 2) a weighted approach that avoids hard thresholds. Weighted gene co-expression networks preserve the continuous nature of gene-gene interactions at the transcriptional level and are robust to parameter selection. An important end product from the gene co-expression network analysis is a set of gene modules in which member genes are more highly correlated with each other than with genes outside a module. Most gene co-expression modules are enriched for GO functional annotations and are informative for identifying the functional components of the network that are associated with disease [75].

This gene co-expression network analysis (GCENA) has been increasingly used to identify gene sub-networks for prioritizing gene targets associated with a variety of common human diseases such as cancer and obesity [38,76-79]. One important end product of GCENA is the construction of gene modules comprised of highly interconnected genes. A number of studies have demonstrated that co-expression network modules are generally enriched for known biological pathways, for genes that are linked to common genetic loci and for genes associated with disease [33,38,71,75-78,80,81]. In this way, one can identify key groups of genes that are perturbed by genetic loci that lead to disease, and that define at the molecular level disease states. Furthermore, these studies have also shown the importance of the hub genes in the modules associated with various phenotypes. For example, GCENA identified ASPM, a hub gene in the cell cycle module, as a molecular target of glioblastoma [78] and MGC4504, a hub gene in the unfolded protein response module, as a target potentially involved in susceptibility to atherosclerosis [77].

Pathway analysis

Resulting lists of differentially expressed genes for each male or female brain area as well as for each module generated in the combined network and some generated in separate networks analysis were analyzed for KEGG (Kyoto Encyclopedia for Genes and Genome, Kyoto University, Japan) pathway enrichment using Pathway-Express, a web-based tool freely available as part of the Onto-Tools (http://vortex.cs.wayne.edu) [82] as well as KEGG website 'Search Pathway' tool (http://www.genome.jp/kegg/tool/search_pathway.html). Global literature analysis of various gene lists was performed using Pathway Studio 8.0 software (Ariadne Genomics, Inc., Rockville, MD).

Additional files

Additional file 1: Movie S1. The first 19 sec of a 10-min mate-preference trial is shown. The trial is conducted under dim red light during the nocturnal (active) phase of the rats' light cycle. At the beginning of the video, the male is in the center of the chamber. The chamber is demarcated into thirds by tape on its floor. A stimulus female can be seen at the far end of the apparatus; the other stimulus female is not visible due to the position of camera. The stimulus females are free-moving in their chambers, but they are separated from the male by a wire mesh that is bounded by Plexiglas barrier. This enables the animals to communicate by olfactory, pheromonal, or behavioral cues, but physical interaction is limited to touching across the wire mesh. The trial begins with the removal of a holding box that confines the male. The male can be seen moving into the zone in front of one stimulus female and then moving across the central portion of the cage to the other stimulus female (out of sight). Several behaviors of the male can be seen on the video such as sniffing, facial investigation, walking, and standing of the female. The male is also seen investigating the various parts of the chamber, including the wire mesh, surrounding Plexiglas partition, and the glass walls of the chamber. Behaviors were scored for each male toward each pair of opposite lineage (Control- or Vinclozolin-Lineage) stimulus females.

Additional file 2: Table S1. Behavior and Sample Information.

Additional file 3: Figure S1A. Samples Histograms After Pre-processing (Male). **Figure S1B.** Samples Histograms After Pre-Processing (Female). Figure S1. Sample histograms and box plots for male (S1A) female (S1B) microarray signal values after pre-processing with RMA, GCcontent adjusted algorithm. Plots for F3 generation control (red) and F3 generation vinclozolin (blue) chips for female amygdala (A), cingulate cortex (B), enterorhinal cortex (C), hippocampus (D), olfactory bulbs (E), and preoptic areaanterior hypothalamus (F).

Additional file 4: Table S2. Genes Differentially Expressed in F3 Generation Vinclozolin Versus Control Lineage Male and Female Rat Brain Regions.

Additional file 5: Top cellular pathways affected by signature gene lists and chosen modules from separate networks.

Additional file 6: Table S3. Pathways Affected Male and Female Brain Region Signature Gene Lists and Chosen Modules from Separate Networks.

Additional file 7: Table S4. Correlation between combined network modules and behavior trait for F3-Vinclozolin rat brain regions.

Additional file 8: Figure S2. (Color) Brain Region Specific Signature List Direct Connection Gene Sub-Networks. Legend: Figure S2. Direct connection sub-networks for signature lists:female amygdala (A), female preoptic area-anterior hypothalamus(B), female hippocampus (C), female enterorhinal cortex (D), female cingulate cortex(E), female olfactory bulbs (F), male amygdala (G), male hippocampus (H), male cingulate cortex(I), male enterorhinal cortex (J), male olfactory bulbs (K) obtained by global literature analysis using Pathway Studio 8.0 software (Ariadne Genomics, Inc., Rockville, MD). Numbers in brackets on figures subtitles indicate number of genes in the list. Only directly connected genes are shown. Some sub-networks (G, H, J) show gene location in the cell (on membrane, in Golgi apparatus, nucleus, cytoplasm or outside the cell). Node shapes and color code: oval and circle – protein; diamond – ligand; circle/oval on tripod platform – transcription factor; ice cream cone – receptor; crescent – kinase or protein kinase; irregular polygon – phosphatase; red color indicates up-regulated genes, blue – down-regulated. Arrows with plus sign show positive regulation/activation, arrows with minus sign – negative regulation/inhibition; grey arrows represent regulation, lilac - expression, purple – binding, green – promoter binding, and yellow – protein modification.

Additional file 9: Table S5. Correlation between separate network modules and behavior trait for F3-Vinlozolin rat brain regions.

Abbreviations

DMR: Differential DNA methylation regions; E8-E14: Embryonic day 8–14; CngCtx: Cingulate cortex; OlfB: Olfactory bulb; k.in: Connectivity index; Amy: Amygdala; EnCtx: Entorhinal cortex; Hipp: Hippocampus; POAH: Preoptic area-anterior hypothalamus; BPA: Bisphenol A.

Competing interests

The authors declare no conflicts of interest.

Authors' contributions

MKS designed the study; MS BZ ACG DC performed the experiments; MKS and DC wrote the paper; MKS MS BZ ACG DC edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Dr. M. Manikkam and Ms. R. Tracey for technical assistance and H. Johnson for assistance in preparation of the manuscript. We thank Drs. C. Guerrero-Bosagna and E. Nilsson for critical review of the manuscript. Research was supported in part by NIH grants ES 017538 (Crews) and ES 012974 (Skinner).

Author details

[1]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA. [2]Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, NY 10029, USA. [3]Pharmacology and Toxicology, Austin, Texas. [4]Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA.

Received: 24 October 2013 Accepted: 28 April 2014
Published: 16 May 2014

References

1. Charlesworth B, Charlesworth D: **Darwin and genetics.** *Genetics* 2009, **183**(3):757–766.
2. Stein RA: **Epigenetics and environmental exposures.** *J Epidemiol Community Health* 2012, **66**(1):8–13.
3. Kalinowski ST: **Evolutionary and statistical properties of three genetic distances.** *Mol Ecol* 2002, **11**(8):1263–1273.
4. Wilkins AS: **Genetic networks as transmitting and amplifying devices for natural genetic tinkering.** *Novartis Found Symp* 2007, **284**:71–86. discussion 86–79, 110–115.
5. Jirtle RL, Skinner MK: **Environmental epigenomics and disease susceptibility.** *Nat Rev Genet* 2007, **8**(4):253–262.
6. Crews D, McLachlan JA: **Epigenetics, evolution, endocrine disruption, health, and disease.** *Endocrinology* 2006, **147**(6 Suppl):S4–S10.
7. Damiani G: **The Yin and Yang of anti-Darwinian epigenetics and Darwinian genetics.** *Riv Biol* 2007, **100**(3):361–402.
8. Day T, Bonduriansky R: **A unified approach to the evolutionary consequences of genetic and nongenetic inheritance.** *Am Nat* 2011, **178**(2):E18–E36.
9. Kuzawa CW, Thayer ZM: **Timescales of human adaptation: the role of epigenetic processes.** *Epigenomics* 2011, **3**(2):221–234.
10. Flatscher R, Frajman B, Schonswetter P, Paun O: **Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation?** *Genet Res Int* 2012, **2012**:698421.

11. Klironomos FD, Berg J, Collins S: **How epigenetic mutations can affect genetic evolution: model and mechanism.** *Bioessays* 2013, **35**(6):571–578.
12. Darwin C: *The Descent of Man, and Selection in Relation to Sex.* London: T. Murray; 1871.
13. Anway MD, Cupp AS, Uzumcu M, Skinner MK: **Epigenetic transgenerational actions of endocrine disruptors and male fertility.** *Science* 2005, **308**(5727):1466–1469.
14. Guerrero-Bosagna C, Settles M, Lucker BJ, Skinner MK: **Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome.** *PLoS One* 2010, **5**(9):e13100.
15. Skinner MK, Manikkam M, Guerrero-Bosagna C: **Epigenetic transgenerational actions of environmental factors in disease etiology.** *Trends Endocrinol Metab* 2010, **21**(4):214–222.
16. Manikkam M, Guerrero-Bosagna C, Tracey R, Haque MM, Skinner MK: **Transgenerational actions of environmental compounds on reproductive disease and epigenetic biomarkers of ancestral exposures.** *PLoS One* 2012, **7**(2):e31901.
17. Skinner MK, Manikkam M, Tracey R, Nilsson E, Haque MM, Guerrero-Bosagna C: **Ancestral DDT exposures promote epigenetic transgenerational inheritance of obesity.** *BMC Med* 2013, **11**:228.
18. Kelce WR, Gray LE, Wilson EM: **Antiandrogens as environmental endocrine disruptors.** *Reprod Fertil Dev* 1998, **10**(1):105–111.
19. Morgan HD, Santos F, Green K, Dean W, Reik W: **Epigenetic reprogramming in mammals.** *Hum Mol Genet* 2005, **14 Spec No 1**:R47–R58.
20. Anway MD, Leathers C, Skinner MK: **Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease.** *Endocrinology* 2006, **147**(12):5515–5523.
21. Anway MD, Skinner MK: **Transgenerational effects of the endocrine disruptor vinclozolin on the prostate transcriptome and adult onset disease.** *Prostate* 2008, **68**(5):517–529.
22. Nilsson EE, Anway MD, Stanfield J, Skinner MK: **Transgenerational epigenetic effects of the endocrine disruptor vinclozolin on pregnancies and female adult onset disease.** *Reproduction* 2008, **135**(5):713–721.
23. Skinner MK, Anway MD, Savenkova MI, Gore AC, Crews D: **Transgenerational epigenetic programming of the brain transcriptome and anxiety behavior.** *PLoS One* 2008, **3**(11):e3745.
24. Anway MD, Rekow SS, Skinner MK: **Transgenerational epigenetic programming of the embryonic testis transcriptome.** *Genomics* 2008, **91**(1):30–40.
25. Skinner MK, Manikkam M, Haque MM, Zhang B, Savenkova M: **Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions.** *Genome Biol* 2012, **13**(10):R91.
26. Crews D, Gillette R, Scarpino SV, Manikkam M, Savenkova MI, Skinner MK: **Epigenetic transgenerational inheritance of altered stress responses.** *Proc Natl Acad Sci USA* 2012, **109**(23):9143–9148.
27. Crews D, Gore AC, Hsu TS, Dangleben NL, Spinetta M, Schallert T, Anway MD, Skinner MK: **Transgenerational epigenetic imprints on mate preference.** *Proc Natl Acad Sci USA* 2007, **104**(14):5942–5946.
28. Friend SH: **The need for precompetitive integrative bionetwork disease model building.** *Clin Pharmacol Ther* 2010, **87**(5):536–539.
29. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusis AJ: **An integrative genomics approach to infer causal associations between gene expression and disease.** *Nat Genet* 2005, **37**:710–717.
30. Yang X, Deignan JL, Qi H, Zhu J, Qian S, Zhong J, Torosyan G, Majid S, Falkard B, Kleinhanz RR, Karlsson J, Castellani LW, Mumick S, Wang K, Xie T, Coon M, Zhang C, Estrada-Smith D, Farber CR, Wang SS, van Nas A, Ghazalpour A, Zhang B, Macneil DJ, Lamb JR, Dipple KM, Reitman ML, Mehrabian M, Lum PY, Schadt EE, *et al*: **Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks.** *Nat Genet* 2009, **41**:415–423.
31. Nilsson EE, Savenkova MI, Schindler R, Zhang B, Schadt EE, Skinner MK: **Gene bionetwork analysis of ovarian primordial follicle development.** *PLoS One* 2010, **5**(7):e11637.
32. Nilsson E, Zhang B, Skinner MK: **Gene bionetworks that regulate ovarian primordial follicle assembly.** *BMC Genomics* 2013, **14**(1):496.
33. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE: **Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks.** *Nat Genet* 2008, **40**(7):854–861.
34. Millstein J, Zhang B, Zhu J, Schadt EE: **Disentangling molecular relationships with a causal inference test.** *BMC Genet* 2009, **10**:23.
35. Pandey G, Zhang B, Chang AN, Myers CL, Zhu J, Kumar V, Schadt EE: **An integrative multi-network and multi-classifier approach to predict genetic interactions.** *PLoS Comput Biol* 2010, **6**(9):e1000928.
36. DiBenedictis BT, Ingraham KL, Baum MJ, Cherry JA: **Disruption of urinary odor preference and lordosis behavior in female mice given lesions of the medial amygdala.** *Physiol Behav* 2012, **105**(2):554–559.
37. Anway MD, Memon MA, Uzumcu M, Skinner MK: **Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis.** *J Androl* 2006, **27**(6):868–879.
38. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S: **Integrating genetic and network analysis to characterize genes related to mouse weight.** *PLoS Genet* 2006, **2**:e130.
39. Sakata JT, Crews D: **Developmental sculpting of social phenotype and plasticity.** *Neurosci Biobehav Rev* 2004, **28**(2):95–112.
40. Skinner MK: **Role of epigenetics in developmental biology and transgenerational inheritance.** *Birth Defects Res C Embryo Today* 2011, **93**(1):51–55.
41. Skinner MK: **Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability.** *Epigenetics* 2011, **6**(7):838–842.
42. Skinner M, Guerrero-Bosagna C, Haque MM, Nilsson E, Bhandari R, McCarrey J: **Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and subsequent germline.** *PLoS One* 2013, **8**(7):e66318.
43. Salian S, Doshi T, Vanage G: **Impairment in protein expression profile of testicular steroid receptor coregulators in male rat offspring perinatally exposed to Bisphenol A.** *Life Sci* 2009, **85**(1–2):11–18.
44. Bruner-Tran KL, Osteen KG: **Developmental exposure to TCDD reduces fertility and negatively affects pregnancy outcomes across multiple generations.** *Reprod Toxicol* 2011, **31**(3):344–350.
45. Waterland RA, Travisano M, Tahiliani KG, Rached MT, Mirza S: **Methyl donor supplementation prevents transgenerational amplification of obesity.** *Int J Obes (Lond)* 2008, **32**(9):1373–1379.
46. Painter RC, Osmond C, Gluckman P, Hanson M, Phillips DI, Roseboom TJ: **Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life.** *BJOG* 2008, **115**(10):1243–1249.
47. Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjostrom M, Golding J: **Sex-specific, male-line transgenerational responses in humans.** *Eur J Hum Genet* 2006, **14**(2):159–166.
48. Dias BG, Ressler KJ: **Parental olfactory experience influences behavior and neural structure in subsequent generations.** *Nat Neurosci* 2014, **17**(1):89–96.
49. Buck LB: **Olfactory receptors and odor coding in mammals.** *Nutr Rev* 2004, **62**(11 Pt 2):S184–S188. discussion S224-141.
50. Brennan PA, Zufall F: **Pheromonal communication in vertebrates.** *Nature* 2006, **444**(7117):308–315.
51. Dulac C, Wagner S: **Genetic analysis of brain circuits underlying pheromone signaling.** *Annu Rev Genet* 2006, **40**:449–467.
52. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ, *et al*: **A draft sequence of the Neandertal genome.** *Science* 2010, **328**:710–722.
53. Kim SY, Kim YS: **A gene sets approach for identifying prognostic gene signatures for outcome prediction.** *BMC Genomics* 2008, **9**:177.
54. Williams G: *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought.* Princeton, NJ: Princeton Univ. Press; 1966.
55. Crews D: **Diversity of hormone-behavior relations in reproductive behavior.** In *Introduction to Behavioral Endocrinology.* 2nd edition. Edited by Becker JBM, McCarthy M, Crews D: MIT Press/Bradford Books; 1992:143–186.
56. Crews D: **The evolutionary antecedents to love.** *Psychoneuroendocrinology* 1998, **23**(8):751–764.
57. Gowaty PA, Anderson WW, Bluhm CK, Drickamer LC, Kim YK, Moore AJ: **The hypothesis of reproductive compensation and its assumptions about mate preferences and offspring viability.** *Proc Natl Acad Sci USA* 2007, **104**(38):15023–15027.
58. Mattle B, Wilson AB: **Body size preferences in the pot-bellied seahorse Hippocampus abdominalis: choosy males and indiscriminate females.** *Behav Ecol Sociobiol* 2009, **63**(10):1403–1410.

59. Carson HL: **The contribution of sexual behavior to Darwinian fitness.** *Behav Genet* 1987, **17**(6):597–611.
60. Carson HL: **Mate choice theory and the mode of selection in sexual populations.** *Proc Natl Acad Sci U S A* 2003, **100**(11):6584–6587.
61. Gowaty PA, Steinichen R, Anderson WW: **Mutual interest between the sexes and reproductive success in Drosophila pseudoobscura.** *Evolution* 2002, **56**(12):2537–2540.
62. Stunden CE, Bluhm CK, Cheng KM, Rajamahendran R: **Factors affecting reproductive performance in captive Mallard ducks.** *Theriogenology* 1999, **52**(3):435–446.
63. Drickamer LC, Gowaty PA, Holmes CM: **Free female mate choice in house mice affects reproductive success and offspring viability and performance.** *Anim Behav* 2000, **59**(2):371–378.
64. Drickamer LCGP, Wagner DM: **Free mutual mate preferences in house mice affect reproductive success and offspring performance.** *Animal Behav* 2003, **65**:105–114.
65. Kirkpatrick M, Ryan MJ: **The evolution of mating preferences and the paradox of the lek.** *Nature* 1991, **350**:33–38.
66. Ryan M: **Sexual selection, sensory systems, and sensory exploitation.** *Oxford Survey Evol Biol* 1990, **7**:157–196.
67. Beach F: **Animal models for human sexuality.** In *Sex, Hormones and Behaviour.* Edited by Potter RWJ. Amsterdam: Ciba Foundation Symposium 62, Excerpta Medica; 1979:113–143.
68. Crews D: **Evolution of neuroendocrine mechanisms that regulate sexual behavior.** *Trends Endocrinol Metab* 2005, **16**(8):354–361.
69. Paxinos G, Watson C: *The Rat Brain in Stereotaxic Coordinates.* New York: Academic; 2007.
70. Bosotti R, Locatelli G, Healy S, Scacheri E, Sartori L, Mercurio C, Calogero R, Isacchi A: **Cross platform microarray analysis for robust identification of differentially expressed genes.** *BMC Bioinforma* 2007, **8**(Suppl 1):S5.
71. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4**:. Article17.
72. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE: **Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations.** *PLoS Comput Biol* 2007, **3**(4):e69.
73. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: **Hierarchical organization of modularity in metabolic networks.** *Science* 2002, **297**(5586):1551–1555.
74. Langfelder P, Zhang B, Horvath S: **Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r.** *Bioinformatics* 2008, **24**(5):719–720.
75. Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, Wang S, Drake TA, Lusis AJ, Schadt EE: **Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes.** *J Neurochem* 2006, **97**(Suppl 1):50–62.
76. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE: **Variations in DNA elucidate molecular networks that cause disease.** *Nature* 2008, **452**:429–435.
77. Gargalovic PS, Imura M, Zhang B, Gharavi NM, Clark MJ, Pagnon J, Yang WP, He A, Truong A, Patel S, Nelson SF, Horvath S, Berliner JA, Kirchgessner TG, Lusis AJ: **Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids.** *Proc Natl Acad Sci U S A* 2006, **103**:12741–12746.
78. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liau LM, Wu H, Geschwind DH, Febbo PG, Kornblum HI, Cloughesy TF, Nelson SF, Mischel PS: **Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target.** *Proc Natl Acad Sci U S A* 2006, **103**:17402–17407.
79. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher, *et al*: **Genetics of gene expression and its effect on disease.** *Nature* 2008, **452**:423–428.
80. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, *et al*: **Mapping the genetic architecture of gene expression in human liver.** *PLoS Biol* 2008, **6**:e107.
81. Zhu J, Zhang B, Schadt EE: **A systems biology approach to drug discovery.** *Adv Genet* 2008, **60**:603–635.
82. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Georgescu C, Romero R: **A systems biology approach for pathway level analysis.** *Genome Res* 2007, **17**(10):1537–1545.

doi:10.1186/1471-2164-15-377
Cite this article as: Skinner *et al.*: Gene bionetworks involved in the epigenetic transgenerational inheritance of altered mate preference: environmental epigenetics and evolutionary biology. *BMC Genomics* 2014 **15**:377.

Genome-Wide DNA Methylation Profiling Reveals Epigenetic Adaptation of Stickleback to Marine and Freshwater Conditions

Artem V. Artemov,*[,1] Nikolai S. Mugue,[2] Sergey M. Rastorguev,[3] Svetlana Zhenilo,[1] Alexander M. Mazur,[1] Svetlana V. Tsygankova,[3] Eugenia S. Boulygina,[3] Daria Kaplun,[1] Artem V. Nedoluzhko,[3] Yulia A. Medvedeva,[1,4] and Egor B. Prokhortchouk*[,1]

[1]Department of Vertebrate Genomics and Epigenomics, Institute of Bioengineering, Research Center of Biotechnology RAS, Moscow, Russia

[2]Russian Federal Research Institute of Fisheries and Oceanography, Moscow, Russia

[3]National Research Center 'Kurchatov Institute', Moscow, Russia

[4]Department of Computational Biology, Vavilov Institute of General Genetics, RAS, Moscow, Russia

***Corresponding authors:** E-mails: artem.v.artemov@gmail.com; prokhortchouk@gmail.com.

Associate editor: Connie Mulligan

Abstract

The three-spined stickleback (*Gasterosteus aculeatus*) represents a convenient model to study microevolution—adaptation to a freshwater environment. Although genetic adaptations to freshwater environments are well-studied, epigenetic adaptations have attracted little attention. In this work, we investigated the role of DNA methylation in the adaptation of the marine stickleback population to freshwater conditions. DNA methylation profiling was performed in marine and freshwater populations of sticklebacks, as well as in marine sticklebacks placed into a freshwater environment and freshwater sticklebacks placed into seawater. We showed that the DNA methylation profile after placing a marine stickleback into fresh water partially converged to that of a freshwater stickleback. For six genes including ATP4A ion pump and NELL1, believed to be involved in skeletal ossification, we demonstrated similar changes in DNA methylation in both evolutionary and short-term adaptation. This suggested that an immediate epigenetic response to freshwater conditions can be maintained in freshwater population. Interestingly, we observed enhanced epigenetic plasticity in freshwater sticklebacks that may serve as a compensatory regulatory mechanism for the lack of genetic variation in the freshwater population. For the first time, we demonstrated that genes encoding ion channels KCND3, CACNA1FB, and ATP4A were differentially methylated between the marine and the freshwater populations. Other genes encoding ion channels were previously reported to be under selection in freshwater populations. Nevertheless, the genes that harbor genetic and epigenetic changes were not the same, suggesting that epigenetic adaptation is a complementary mechanism to selection of genetic variants favorable for freshwater environment.

Key words: **epigenetics, DNA methylation, stickleback, adaptation, evolution, water salinity, marine, freshwater, fish, phenotypic variation, ion pumps**

Introduction

The three-spined stickleback (*Gasterosteus aculeatus*) is an important model in the study of adaptive evolution. The species can survive in environments with a large variety of salinity, such as river estuaries and the brackish waters of the Baltic Sea (McCairns and Bernatchez 2010; Guo et al. 2015; Konijnendijk et al. 2015). Successful colonization of different environmental conditions involves rapid adaptation to factors such as a sharp change in temperature, salinity, other predators, and parasites (Barrett et al. 2011; Lescak et al. 2015).

Isolated freshwater populations of sticklebacks are believed to originate by separation from the marine population. Many independent acts of river and lake colonization by marine sticklebacks have been reported (Jones et al. 2012), making it possible to study a wide variety of evolutionary trajectories of adaptation to freshwater. Parallel adaptation between independently formed freshwater populations to a new habitat occurs by increasing the frequency of certain freshwater alleles that preexist at low frequency in the marine population (Hohenlohe et al. 2012; Jones et al. 2012). Adaptation to heterogeneous environmental conditions, such as the Baltic Sea, and to the freshwater lakes and creeks have similar genomic mechanisms (Roesti et al. 2014; Guo et al. 2015). They operate within genomic loci that contain mainly regulatory sequences rather than protein-coding regions (Jones et al. 2012). The plasticity of gene expression in response to changing environmental conditions is likely to help the three-spined stickleback colonize a wide range of habitats (McCairns and Bernatchez 2010; Morris et al. 2014).

High salt conditions represent a major challenge for living organisms because they cause DNA double-strand breaks and cell senescence (Dmitrieva and Burg 2007; Dmitrieva et al. 2011).

Article

It has been reported that marine invertebrates have numerous DNA breaks due to high salt conditions (Dmitrieva et al. 2006). It is therefore of fundamental interest to explore the mechanisms of how organisms adapt to different osmotic conditions.

Several studies revealed genetic aspects of *Gasterosteus aculeatus* freshwater adaptation. Terekhanova and colleagues (Terekhanova et al. 2014) defined divergence islands as regions with a significant shift of allele frequency between marine and freshwater populations, which are presumably under selection in freshwater populations. Some divergence islands harbored genes that are believed to be associated with freshwater adaptation, such as ion transporters. Notably, some of the divergence islands lacked genes, suggesting their regulatory potential as enhancers. Reduction of the pelvic apparatus characteristic to many freshwater populations was shown to be caused by the loss of a tissue-specific enhancer of the Pitx1 gene (Chan et al. 2010). Among the genomic loci, associated with freshwater adaptation reported in (Jones et al. 2012), noncoding regulatory changes appeared to be predominant.

Even though genetic components are important to adaptation to osmotic conditions, various organisms can switch between sea and freshwater habitats within one generation, suggesting that epigenetic mechanisms, such as histone modifications, regulation by microRNA (Rastorguev et al. 2016), and DNA methylation (Varriale 2014) might be involved in adaptation. There are several examples when variations in phenotypic traits are accompanied by minimum genetic diversity, but a significant level of epigenetic diversity. Noteworthy examples include bats (Liu et al. 2012, 2015) and plants (Gao et al. 2010; Yi et al. 2010). A study in Arctic charr (*Salvelinus alpinus*) revealed that seawater exposure induced changes in DNA methylation and peroxynitrite formation in gills (Norman et al. 2014). In a study aimed to understand the functionality of miRNA in stickleback adaptation to freshwater environments, ten miRNAs were found in divergence islands and two miRNAs contained SNPs with shifted allele frequencies between the freshwater and the marine populations (Rastorguev et al. 2016). DNA methylation was shown to be associated with phenotypic variability between complete and low lateral plate morphs in the freshwater stickleback population (Smith et al. 2015). Discovered differentially methylated regions (DMRs, mostly intergenic) were associated with genes having potentially adaptive functions, including cardiovascular development, growth, and neuromuscular development. However, the study was focused only on freshwater fish and epigenetic adaptation to water salinity has not been studied yet.

High levels of genetic variation within a population may impose a high probability of adaptation of the population to a new environment. However, isolation of a small population inevitably leads to decreased genetic diversity. Yet, to provide a substrate for natural selection in a new environment, a population should demonstrate a certain amount of diversity. Such diversity may be achieved by epigenetic variations, since genetically inherited propensity to phenotypic variability, even with no change in the mean phenotype, substantially increases fitness (Feinberg and Irizarry 2010). The idea of an evolutionary benefit of increased epigenetic variability is formulated as the epigenetic plasticity hypothesis.

In this work, we performed whole-genome DNA methylation profiling of marine and freshwater sticklebacks, as well as sticklebacks moved for 4 days from a marine environment to a freshwater environment and vice versa. We investigated the role of DNA methylation in the short-term (4 days of exposure to a foreign environment) and long-term (differences between marine and freshwater populations) adaptation to changed salinity. We compared the differences in DNA methylation between marine and freshwater populations with genetic adaptations of a freshwater population studied in (Terekhanova et al. 2014). To confirm that changes in DNA methylation have a functional effect, we also investigated the expression of the genes associated with differentially methylated regions in the same sample groups.

Results

Changes in Water Salinity Have an Impact on DNA Methylation in Sticklebacks

To explore the role of DNA methylation in adaptation to freshwater conditions, we studied the following sample groups: i) marine sticklebacks kept in marine water, their natural habitat (M@M); ii) freshwater sticklebacks kept in freshwater (F@F); iii) marine sticklebacks incubated in fresh water for 4 days (M@F); iv) freshwater sticklebacks incubated in marine water for 4 days (F@M), (fig. 1).

DNA methylation was profiled in gills because this organ comes into direct contact with the surrounding water and is likely to be highly affected by osmotic stress. Three comparisons were performed between the studied groups of individuals to discover differential methylation between marine and freshwater populations of sticklebacks in water with natural salinity (M@M vs. F@F), methylation changes induced by placing a marine stickleback into a freshwater environment (M@M vs. M@F) and methylation changes induced by placing a freshwater stickleback into a marine environment (F@F vs. F@M). We found 61, 245 and 26 DMRs for the three listed comparisons (M@M vs. F@F, M@M vs. M@F, F@F vs. F@M, respectively), (Supplementary fig. S1A, Supplementary Material online). Two DMRs were validated with bisulfite treatment followed by Sanger sequencing (Supplementary fig. S2, Supplementary Material online). The number of DMRs discovered in each comparison did not reflect the difference between groups: MM and MF groups were the most similar but had the highest amount of DMRs if compared against each other (Supplementary fig. S3, Supplementary Material online). Instead, more DMRs were discovered between groups having low within-group variance.

Most of the DMRs we discovered were located within annotated genes or within 1 kb from gene boundaries (Supplementary fig. S4, Supplementary Material online), even though genes and gene flanks occupy no more than a quarter of the genome. Nevertheless, the observed significant

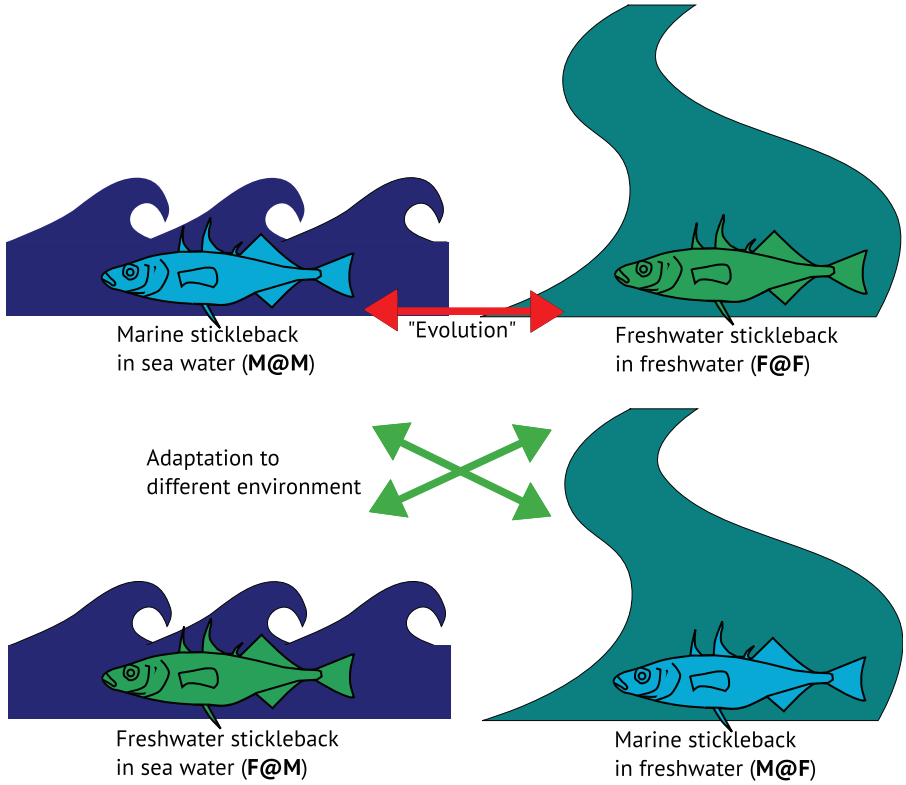

FIG. 1. Experimental design. DNA methylation was profiled in four conditions: marine stickleback in marine and freshwater environment, freshwater stickleback in freshwater and marine environment. Differentially methylated regions were found between marine and freshwater sticklebacks in their respective natural habitats. These methylation changes were compared with the immediate changes caused by placing marine sticklebacks into a freshwater environment. Additionally, a similar comparison was made to find DNA methylation changes after placing freshwater sticklebacks into a sea environment.

enrichment of DMRs within genes (p(Binomial) between $2*10^{-30}$ and 10^{-5}) can be explained by increased CpG density in the genic regions and in the regions profiled by RRBS rather than by the functional role of DMRs. We associated the DMRs with genes (see Materials and Methods). We called a gene differentially methylated (DMG) if it was linked to at least one DMR. We detected 40, 151 and 16 DMGs for M@M vs. F@F, M@M vs. M@F, and F@F vs. F@M comparisons, respectively (Supplementary fig. S1*B*, Supplementary Material online).

We also validated functional changes in expression of the genes associated with DMRs by RNA-seq. Out of 11 DMGs that were differentially expressed between marine and freshwater populations (M@M vs. F@F), eight genes had expression changes in line with the direction of change in DNA methylation: seven genes were hypomethylated and upregulated, one gene was hypermethylated and downregulated. This result indicated that, in general, DNA methylation was associated with repressed transcription.

Some of the observed DMRs were located in gene bodies which can be explained by increased GC content in exons and the bias of RRBS towards GC- and CpG-rich sequences. DMRs within gene bodies can occur at intergenic enhancers or at alternative promoters—in this case we would expect an increase of gene expression following hypomethylation of a DMR. Alternatively, DNA methylation can be changed because of altered transcription elongation in gene body—in this case increased expression is expected to cause increase in DNA methylation. Therefore, we observe negative correlation between DNA methylation and expression only in some of the detected DMGs.

Ion Channels, Membrane Proteins and Regulatory Genes Are Differentially Methylated between Marine and Freshwater Sticklebacks

To investigate long term evolutionary adaptation, we compared M@M to F@F (MMFF). A list of DMGs for MMFF consisted of 40 genes (Supplementary fig. S1*B*, Supplementary Material online). Gene category enrichment analysis revealed that the genes differentially methylated in the freshwater stickleback (F@F) population, as compared with the marine (M@M) population, were significantly enriched for ion channels and transmembrane proteins (fig. 2*A*). In fact, all the categories discovered were associated with membrane-bound proteins. The discovered ion channels included CACNA1FB (ENSGACG00000000800)—calcium channel, voltage-dependent, L type, alpha 1F subunit, H+/K+ exchanging ATPase ATP4A (ENSGACG00000008911) and potassium voltage-gated channel KCND3 (ENSGACG00000000195), as well as gap junction protein GJA3 (ENSGACG00000001367). In figure 3*A*, we summarized the methylation changes in individual CpGs in the DMR associated with one of these ion channels, CACNA1FB. A coordinated decrease in methylation was observed in each CpG within the DMR in the freshwater population (FF) compared

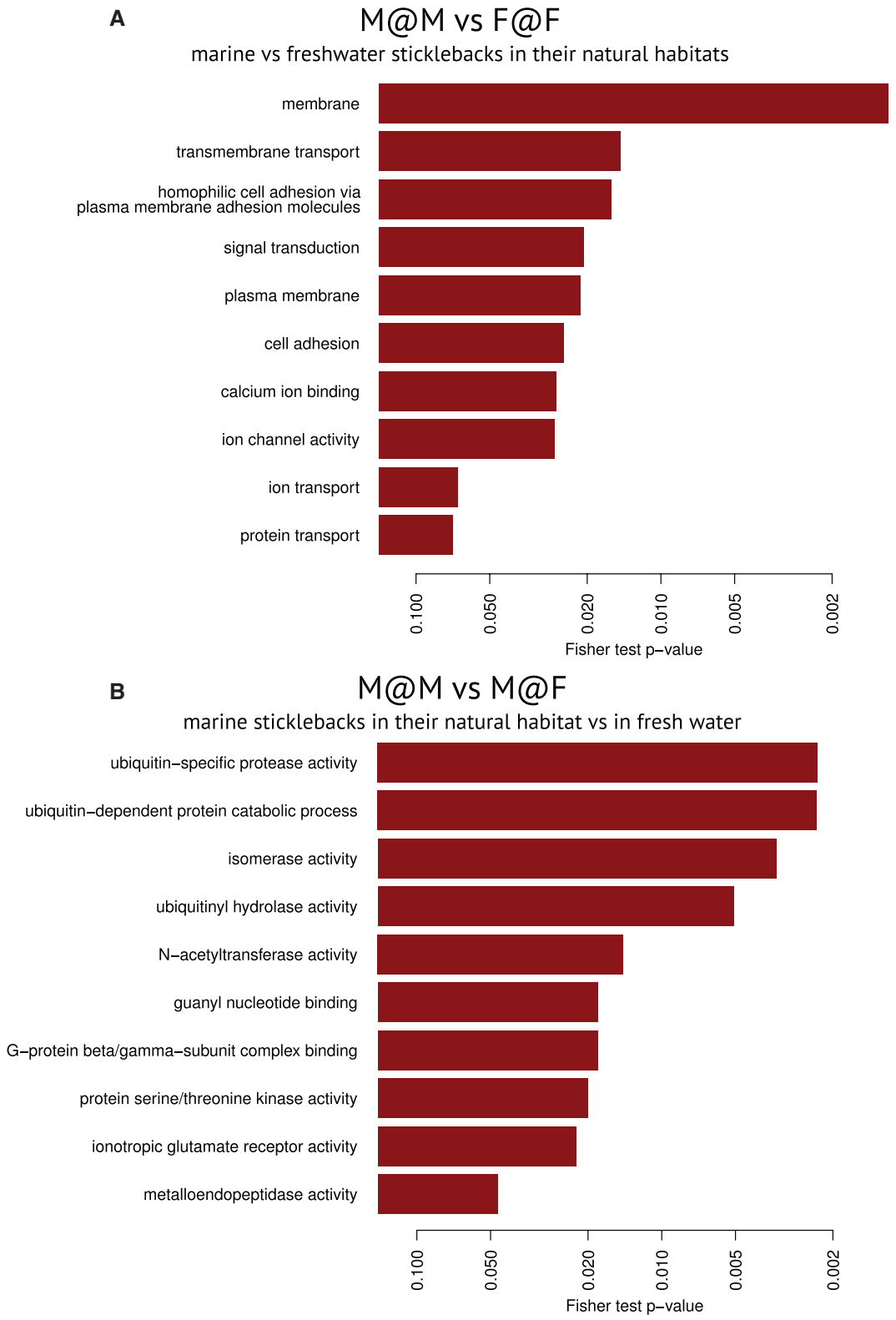

Fig. 2. GO categories enrichment plot for genes associated with DMRs between (*A*) marine and freshwater populations; (*B*) marine population in its native environment and marine sticklebacks placed into fresh water. *X*-axis shows enrichment *P* value according to Fisher exact test (logarithmic scale). Only categories with more than one differentially methylated gene were considered.

with the marine population (MM). Moreover, we showed that expression of CACNA1FB was significantly higher in the freshwater population than the marine population (fig. 3*B*, *P* value of differential expression between MM and PP groups $3.9*10^{-10}$, FDR $7.9*10^{-8}$). The direction of changes in DNA methylation and gene expression were in line with a common concept that DNA methylation in gene promoters was negatively correlated with gene expression. Similar figures

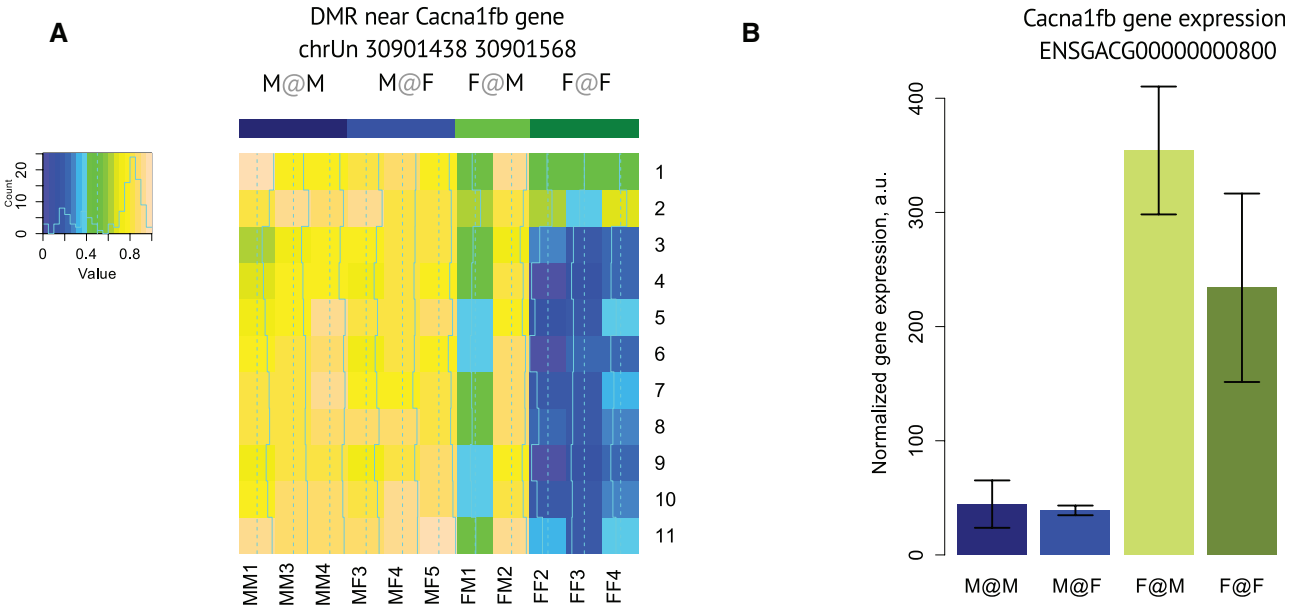

Fig. 3. DNA methylation status and gene expression of CACNA1FB (ENSGACG00000000800) gene in the profiled samples. Figure (*A*) shows methylation values of individual CpGs within a DMR found near the gene. Only positions covered at least ten times in each sample were considered. Figure (*B*) shows normalized expression counts for the gene derived from RNA-seq data. DESeq test of differential expression between MM and PP groups yielded a significant *P* value of $3.9*10^{-10}$, even after multiple testing correction (FDR = $7.9*10^{-8}$).

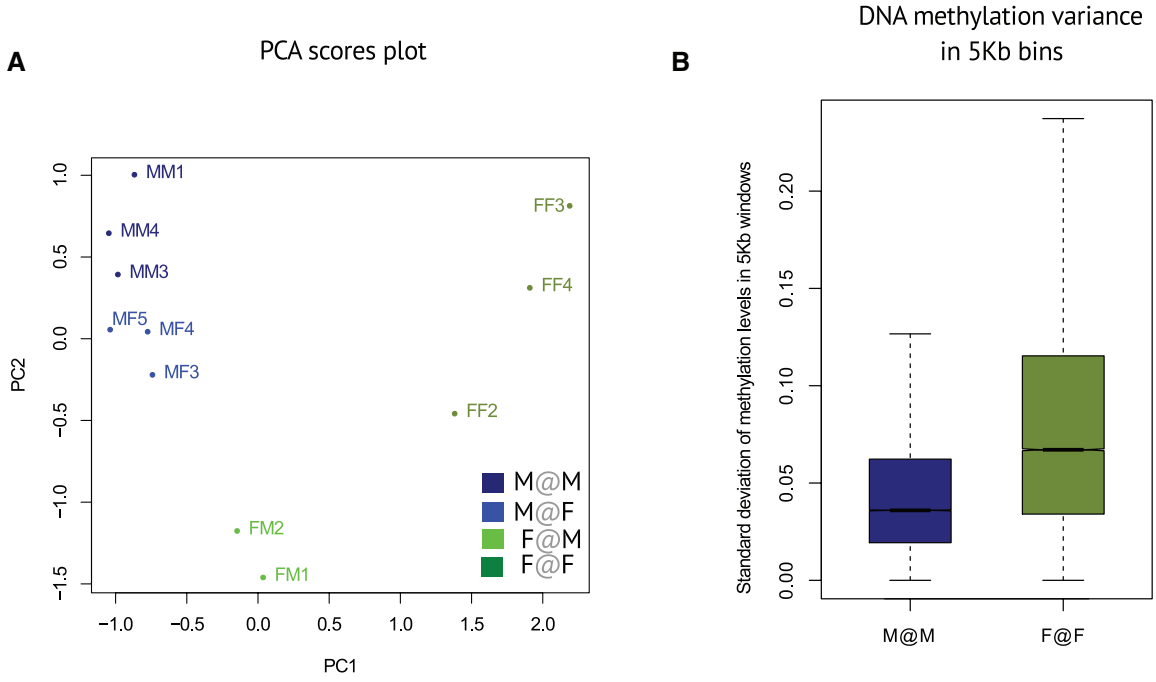

Fig. 4. (*A*) PCA plot showing all studied samples in four experimental groups. First principal component explained 42% of total variance, first two principal components explained together 59% of total variance. (*B*) Variance in DNA methylation levels in 5 kb genomic bins between individuals from stickleback populations. Significantly higher variance was observed in the freshwater population compared with the marine population, which could compensate for the lack of genetic variance in the freshwater population.

for other DMGs are provided in Supplementary Text S2, Supplementary Material online.

It is widely accepted that in evolution, marine sticklebacks colonized freshwater environments (and not the other way around), (Jones et al. 2012; Terekhanova et al. 2014). Therefore, we wanted to determine if long-term evolutionary adaptation to freshwater involved the same mechanisms as the short-term adaptation that could be observed in marine stickleback put into fresh water (M@F). As expected, correlation of methylation profiles of the studied MM, FF, and MF samples suggested that MF samples were more similar to their original marine population (MM) than to the freshwater population (Supplementary fig. S3, Supplementary Material online). Similar clustering was observed for RNA-seq data

(Supplementary fig. S5, Supplementary Material online). However, six genes were associated with significant methylation changes both between freshwater and marine populations (F@F|M@M) and in marine sticklebacks placed into freshwater compared with marine populations (M@F|M@M) (Supplementary fig. S1*B*, Supplementary Material online). Among them, we discovered an ion pump, alpha polypeptide of H+/K+ exchanging ATPase ATP4A (ENSGACG00000008911). Other genes were likely to have regulatory function: sirtuin SIRT2, a deacetylase involved in many cellular processes including histone deacetylation (ENSGACG00000005747), protein kinase C-binding protein NELL1 (ENSGACG00000017098), ubiquitin specific peptidase USP20 (ENSGACG00000015949), methylthioadenosine phosphorylase MTAP (ENSGACG00000017558), and M-phase phosphoprotein MPHOSPH9 (ENSGACG00000015639). Interestingly, NELL1 is believed to control skeletal ossification (James et al. 2015) and therefore may partially contribute to the well-known difference in the phenotypes of armor plates between marine and freshwater sticklebacks (Bell 2001).

DNA Methylation After a Short-Term Adaptation to Freshwater Partially Mimics Evolutionary Adaptation

As we have shown, for some genes, including the ion pump ATP4A, DNA methylation change induced by environmental stress mimicked that between marine and freshwater populations, but it was still unclear how universal this effect could be. First, we explored the overlap between the sets of DMGs from each of the three comparisons: M@M vs. F@F, M@M vs. M@F, and F@F vs. F@M (Supplementary fig. S1*B*, Supplementary Material online). The overlaps of 6 DMGs shared between the M@M vs. F@F and M@M vs. M@F comparisons and 3 DMGs shared between the M@M vs. F@F and F@F vs. F@M comparisons seemed relatively small, yet significant considering total number of genes not associated with DMRs (P values of $4*10^{-19}$ and 0.018, respectively; see Materials and Methods for the statistical procedure).

To figure out how similar immediate adaptations were to interpopulation differences on the genome scale, we performed a PCA analysis of DNA methylation levels in 5 kb genomic windows in all studied samples (fig. 4*A*). The samples fall into four groups according to their origin. A possible interpretation of the PCA plot could associate PC1 with adaptation to freshwater conditions (PC1 discriminates best between marine and freshwater populations in their natural habitats and F@M samples fall in between), whereas PC2 might reflect a general stress response after placing a fish into a different environment. Enrichment of the genes differentially methylated after placing fish into a different environment, for example, M@M vs. M@F, within gene categories related to signal transduction, particularly, ubiquitination, G-protein signaling, and kinases (fig. 2*B*), supported the idea that PC2, which separated fish in their native and foreign environments, was related to stress response. Taken together, our results suggested that some of the prospective epigenetic adaptations to decreased water salinity emerged both in the freshwater population and in marine sticklebacks "*de novo*" placed into freshwater. In other words, some of the immediate epigenetic adaptations to freshwater could be maintained in populations over time.

The Interplay between Genetic and Epigenetic Adaptation

To find out if DNA methylation can play a role in epigenetic activation or suppression of genes or regulatory regions that were under selection in the freshwater population (divergence islands, DI, [Terekhanova et al. 2014]), we searched for DMRs within DIs. We discovered that DI XXI-1 (chrXXI 5757849 7491073) overlapped with 3 DMRs from the M@F-M@M comparison, and no other overlaps were found (Supplementary fig. S6, Supplementary Material online).

As DNA methylation was known to be a major factor influencing mutagenesis of cytosines (C to T transitions), we next studied if DMRs were enriched by rare polymorphisms with alternative alleles present only in freshwater populations and absent in the marine population. The DMRs overlapped with neither rare freshwater-specific polymorphisms, nor any other SNPs having significantly different allele frequencies between marine and freshwater populations. The obtained results are not surprising, as the studied set of SNPs was believed to contain mostly SNPs that were preexisting in the marine population and selected in freshwater populations, rather than *de novo* mutations happened in freshwater populations.

Variation in DNA Methylation Is Increased in the Freshwater Population

We hypothesized that for the freshwater stickleback population, which had passed a population bottleneck and thus had limited genetic variability, it would be beneficial to compensate for the lack of heterozygosity by increased epigenetic variance. We calculated the variance of DNA methylation in 5 kb consecutive bins throughout the whole genome in the original marine (M@M) and freshwater (F@F) populations. Indeed, the freshwater population (F@F) demonstrated significantly higher variance in DNA methylation than the marine population (M@M, Wilcoxon test P value $< 2.2*10^{-16}$), suggesting that DNA methylation may contribute to freshwater adaptation by increasing the phenotypic variance (fig. 4*B*, Supplementary fig. S7, Supplementary Material online). This would yield a wider window of possibilities for a population to adapt to environmental challenges.

Discussion

Adaptation of marine sticklebacks to freshwater conditions was considered as a model system to study underlying genetic events. Freshwater fish were collected from Mashinnoe lake, which was physically separated from the sea by the motion of tectonic plates around 700 years ago (Kolka and Korsakova 2005). From this time on, the salinity of the lake dropped to a freshwater level. Current understanding of adaptation includes a model where preexisting alleles in the seawater population have selective advantages that allow certain fish to survive in freshwater environments. There was a number of studies that resolved genetic mechanisms and pointed to the

alleles that change their frequency upon this adaptation (Jones et al. 2012; Terekhanova et al. 2014). However, little is known on how the stickleback epigenome responds to altered salinity. We aimed to detect changes in DNA methylation in fish living at different salt concentrations.

Similarly to previously discovered selection-driven genetic alterations between marine and freshwater populations, the epigenetic changes which we found in this study were associated with transmembrane proteins and particularly ion pumps. Differentially methylated genes included a calcium channel CACNA1FB, a potassium channel KCND3, an H+/K+ exchanging ATPase ATP4A and a gap junction protein GJA3. Strikingly, genetic studies of marine and freshwater sticklebacks discovered functionally similar genes, such as potassium channel KCNH4 (Jones et al. 2012) and an ion-exchanging ATPase ATP1A1 (Terekhanova et al. 2014). Nevertheless, the sets of genes that were affected genetically and epigenetically did not overlap. This fact supported the idea that genetic and epigenetic adaptations represent independent mechanisms, which are both necessary for environmental adaptation. Moreover, it suggested that the observed epigenetic differences did not arise solely *in cis* due to different genetic backgrounds of fish populations. However, our model does not exclude a possibility that DNA methylation is affected by genetic background *in trans*.

Remarkably, the differences in DNA methylation between the marine and the freshwater populations had a significant similarity to the changes in DNA methylation that were induced by short-term exposure of marine fish to freshwater environment. A significant overlap of the DMGs found in both of those two comparisons included an ion pump (ATP4A) and several regulatory genes. Differential methylation in one of these genes, NELL1, could potentially affect skeletal ossification and explain the difference in the phenotypes of armor plates between marine and freshwater sticklebacks (Bell 2001; James et al. 2015). Alteration of DNA methylation in the genes that were differentially methylated in both comparisons might represent an initial response to freshwater conditions, which cannot be substituted by other evolutionary changes and thus the difference remains between the populations. However, the current data are not sufficient to distinguish if the methylation pattern characteristic to marine or freshwater populations is inherited through meiosis or results from permanent exposure to freshwater or marine environment.

The genes that were differentially methylated after placing a marine stickleback into freshwater were enriched with stress response and signal transduction genes, which resembled the response of other organisms to osmotic stress. In particular, we observed differential methylation of the genes involved in ubiquitination and autophagy, such as ubiquitin-specific peptidase USP16, ubiquitin carboxyl-terminal esterase L3 UCHL3, ubiquitin conjugating enzyme E2 E1 UBE2E1, as well as mitogen-activated protein kinase kinase kinase 4 (MAP3K4) which is believed to be activated in response to stress, including osmotic stress. Previously, osmotic stress was shown to cause accumulation of polyubiquitinated proteins through activation of p38 MAPK kinases which in turn inhibited proteasome activity (Whitmarsh and Davis 2007; Lee et al. 2010). Hypermethylation of MAP3K4 kinase and hypomethylation of uchl3 observed in marine sticklebacks placed into freshwater could in principle reduce protein ubiquitination level and serve as an adaptation to osmotic stress.

DNA methylation can potentially contribute to gene inactivation in two ways: on one hand, *de novo* methylated regions mark genes with altered expression, on the other hand, methylated cytosines are targets for increased (up to 5-fold) cytosine-uracil-thymine mutability through deamination (Cooper and Krawczak 1989). On a population level, DNA methylation was also shown to greatly influence cytosine mutation rate in humans (Xia et al. 2012). Thus, changes in DNA methylation in germline cells presumably had a potential to become fixed in the genome upon sea to freshwater adaptation, which could theoretically lead to increased abundance of C > T transitions between sea (C) and freshwater (T) sticklebacks among C's that are *de novo* methylated in freshwater fish. However, we found no overlap between polymorphisms that have differential allele frequencies in marine and freshwater sticklebacks and DMRs in any of the three comparisons. This observation does not disprove the hypothesis that DNA methylation could facilitate genetic silencing of genes by increasing the mutation rate of methylated CpGs in their regulatory regions, as the studied polymorphisms reflected selection rather than mutagenesis. Due to low sample sizes, current studies are unable to fetch polymorphisms with low minor allele frequencies. Therefore, most mutational effects are likely to be lost.

In Feinberg and Irizarry (2010), a new concept of variably methylated regions (VMRs) was defined as regions in which DNA methylation varies stochastically across the population, even within the same tissue and even in isogenic mice. In a conventional model, marine sticklebacks represent a genetic reservoir that gives rise to a freshwater population with minimal genetic diversity, but with a capability to live both at high and low salt concentrations. We formulated a hypothesis that a population challenged by minimal genetic diversity might adapt to a new environment and maintain a certain level of phenotypic variation by having a high level of epigenetic variability. Hence, we expected to see higher variance of DNA methylation in the freshwater population compared with the marine population. Indeed, figure 4*B* confirms this statement. In an attempt to find a mechanism driving higher epigenetic variance in the freshwater population, we checked whether genomic loci that were reported by Terekhanova et al. (2014) and served as landmarks of the genetic component of adaptation included genes that were responsible for chromatin architecture or DNA methylation. However, we found no chromatin regulators among the genes within the regions under selection. Genetics that could underlie higher epigenetic plasticity in freshwater sticklebacks still remains unclear.

Materials and Methods

Sample Collection

Ten freshwater three-spined sticklebacks were collected in the Mashinnoye Lake, Republic of Karelia, Russia, in

June–July 2014. The same number of marine three-spined sticklebacks were collected in the Kandalaksha Gulf Coast (15 km from the railway station Poyakonda, Republic of Karelia, Russia; water salinity of 26 promille) during the same time frame. To ensure equal physiological states of the fish, only males in breeding dress were collected.

To explore the immediate epigenetic response of a stickleback to changes in osmotic conditions of the environment, five marine fish were kept for 4 days in fresh water (M@F) and, vice versa, five freshwater fish were kept for 4 days in marine water (F@M, water salinity of 26 promille). We chose the exposure time to let the epigenome adapt to marine or freshwater conditions. It has been shown that for a cell line it takes up to 24 h to establish epigenome changes after exposure to high salt conditions (Dmitrieva et al. 2011). We hypothesized that it could take longer for the organism tissue to adapt, therefore we chose 4 days as a period of exposure to a different salinity conditions. It should be noted that control samples (M@M, F@F) were kept in the water with native salinity for the same time period to equal captivity stress influence.

Library Preparation, RRBS Sequencing and Data Processing

Control fish were kept in their native habitats (marine samples in marine water (M@M), freshwater samples in fresh water (F@F)) to equal captivity stress influence. After 4 days of exposure, DNA was isolated from gills (an organ that is directly exposed to water is likely to be highly affected by osmotic stress) using a standard method of DNA extraction from animal tissue (phenol-chloroform). The concentration of DNA in the samples was measured using a Nanodrop spectrophotometer (Supplementary table S2, Supplementary Material online).

Three individual fish from each of the four experimental conditions were taken for bisulfite sequencing (Supplementary table S2, Supplementary Material online). DNA methylation was profiled by reduced representation bisulfite sequencing (RRBS). Two micrograms of genomic DNA from three-spined stickleback samples of *G. aculeatus* of (Supplementary table S2, Supplementary Material online) and 4 ng of lambda phage DNA were digested using 60 U MspI (Fermentas, USA) in 50 µl at 37°C for 18–24 h, followed by QIAquick purification (Qiagen, Germany). The end of the digested DNA was repaired, and an adenine was added to the 3′ end of the DNA fragments according to the Illumina standard end repair and add_A protocol (Illumina, USA). Preannealed forked Illumina adaptors containing 5′-methylcytosine instead of cytosine were ligated to both ends of DNA fragments using standard Illumina adaptor ligation protocol (Illumina, USA). Ligated fragments were then separated by 2% agarose gel (Sigma-Aldrich, USA). Fragments between 170 and 350 bp, (includes adaptor length), were selected and cut from the gel. DNA from gel slices were purified using the Qiagen Gel Extraction Kit (Qiagen, Germany). The sodium bisulfite treatment and subsequent clean-up of size selected DNA was performed with the EZ DNA Methylation™ Kit (ZymoResearch, USA) according to the manufacturer's instructions. The bisulfite-treated DNA fragments were amplified using PCR and the following reaction: 5 µl of eluted DNA, 1 µl of NEB PE PCR two primers (1.0 and 2.0), and 45 µl Platinum PCR Supermix (Invitrogen, USA). The amplification conditions were as follows: 5 min at 95 °C, 30 s at 98 °C then 15 (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C), followed by 5 min at 72 °C. The PCR reaction was purified by MinElute PCR Purification Kit (Qiagen), and final reduced representation bisulfite library was eluted in 15 µl EB buffer. The concentration of the final library was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The library was sequenced on Illumina 2500 platform according to standard Illumina cluster generation and sequencing protocols. 100-bp single-end reads were generated.

Reads were mapped to gasAcu1 reference genome (augmented with the genomic sequence of lambda phage) with Bismark software (Krueger and Andrews 2011). As the spiked-in phage DNA was unmethylated, the methylation level of cytosines in lambda phage DNA profiled in the experiment served as a measure of how incomplete the bisulfite conversion was. The observed methylation levels of 0.30–0.65% for all sample except FM3 suggested that bisulfite conversion was 99.35–99.70% complete (Supplementary table S3, Supplementary Material online, "RRBS" sheet). However, for FM3 sample bisulfite conversion appeared to be incomplete (41%). This sample from the FM experimental group was excluded from the analysis after quality control. Individual samples from each experimental groups were treated as biological replicates. With bsseq R package (Hansen et al. 2012), we found individual differentially methylated CpG's and differentially methylated regions (DMRs). Default parameters for bsseq were used: minimal coverage per CpG 10, *t*-statistics cutoff 4.6, minimal mean methylation difference between groups 0.1, minimal number of CpGs per DMR 3. Individual DMRs were visualized to inspect methylation frequencies and read coverage per individual CpGs in individual samples. RRBS produced data for each CpG dinucleotide, thus giving information with single nucleotide resolution. To enhance the statistical power of the analysis, instead of looking for individual CpG dinucleotides which altered methylation level, we searched for differentially methylated regions (DMRs) that contain at least three CpG and demonstrate at least 10% difference in methylation. ENSEMBL gene annotation (version 81) was used to find genes localized no further than 1 kb from differentially methylated sites.

To estimate the false discovery rate (FDR) of the DMR discovery procedure, we applied a permutation approach. In every iteration, we randomly selected two groups of three samples each and ran the DMR discovery procedure, which would find differentially methylated regions between those two groups. We excluded the cases in which any of the two permuted groups was equal to one of the experimental groups (e.g., all samples in a permuted group came from the MM group). The procedure was repeated 100 times. This procedure yielded a median of 20 DMRs per iteration between permuted groups of samples, which can be used as an expected number of false discoveries. Considering that samples from all experimental groups were used for permutations, we compared the expected number of false

discoveries to the average number of real DMRs discovered in the MMFF and MMMF comparisons. We only considered the comparisons in which a group of three samples was compared with a group of three samples (i.e., MMFF and MMMF) as a lower number of samples in an experimental group (e.g., two samples in the FM group) can affect the number of the discovered DMRs. Based on these considerations, the FDR can be estimated on the level of 0.07 (20 expected false discoveries vs. 153 real hits).

For DMR validation, genomic DNA from gills was purified using a standard procedure of DNA extraction from animal tissue (phenol-chloroform). Bisulfite conversion was performed by EZ DNA methylation Kit (Zymo Research) according to the manufacturer's instructions. After bisulfite conversion, selected regions were amplified by the following pairs of primers:

TTGGATCCGTTTGGTATGTAATTATATTTGGT and

TTGAATTCACTACTACGAAATAAATAACACCC;

TTGGATCCTGATTGAATATGTAGTATGTAGTT and

TTGAATTCAACTCAACTACAACTCAATA.

PCR products were cloned to the T-vector. For each region ten clones were sequenced by Sanger and analyzed. The two profiled regions appeared to be differentially methylated in Sanger sequencing (Fisher's exact test P values were $<2.2*10^{-16}$ and $2.3*10^{-9}$; Supplementary fig. S2, Supplementary Material online). We also used DNA methylation values profiled by Sanger sequencing to validate precision of RRBS profiling for individual CpGs. Overall correlation of methylation rate profiled by RRBS and Sanger for individual CpGs in corresponding sample groups was 0.900 (P value $= 2*10^{-9}$).

To describe how the DMRs were localized with respect to genes and gene flanks, we calculated the number of DMRs within genes, the number of DMRs located no further than 1 kb from a gene and the number of DMRs in intergenic regions. We also estimated the total genomic length of genic regions, gene flanks, and intergenic region. A background probability of a DMR to appear within a gene or a gene flank under a null hypothesis of no enrichment was estimated as a fraction of the total length of genic regions and gene flanks in the genome. We performed a binomial test to estimate if the DMRs were enriched within genes and gene flanks.

We annotated the DMRs with respect to known stickleback genes. In particular, we searched for all differentially methylated regions (DMR) within 1kb upstream and 1kb downstream of all annotated genes. We considered a gene to be differentially methylated (DMG) if it had a DMR no further than 1 kb from it (Supplementary table S1, Supplementary Material online). Additionally, we provide average DNA methylation within the 2 kb regions surrounding a TSS for each transcript (Supplementary table S4, Supplementary Material online).

DMGs were further used for gene category enrichment analysis. Annotation of genes related to GO categories was taken from ENSEMBL database. Gene set enrichment analysis was done as follows: for each category having at least one gene in a gene set, we performed Exact Fisher's test to check if the category is enriched in a given gene set.

To estimate the significance of overlap between two sets of DMRs (e.g., M@M vs. F@F and M@M vs. M@F) under a strict assumption that most of DMRs were associated with gene promoters, the following test was performed. For each gene, we asked if a DMR from each of the two sets was associated with the gene. Next, we performed Fisher exact test to check if presence of a DMR in one set was associated with presence of a DMR in the other set.

RNA Preparation and RNA Sequencing

For RNA-seq, we used the same sample collection and treatment procedure as described for bisulfite sequencing. Four fish from each of the four experimental groups were taken for transcriptome analysis. Gills were isolated and fixed with IntactRNA® reagent (Evrogen).

Total RNA was extracted from the samples with Trisol reagent according to the manufacturer's instructions (Invitrogen). Quality was checked with the BioAnalyser and RNA 6000 Nano Kit (Agilent). PolyA RNA was purified with Dynabeads® mRNA Purification Kit (Ambion). An Illumina library was made from polyA RNA with NEBNext® mRNA Library Prep Reagent Set (NEB) according to the manual. Paired-end sequencing was performed on HiSeq1500 with 2 $\times$ 75 bp read length. Approximately 25 million reads were generated for each sample.

Reads were mapped to gasAcu1 genome with tophat2 software (version 2.1.0) (Kim et al. 2013). Number of RNA-seq reads, number of unmapped reads and mapping efficiency is summarized in Supplementary table S3, Supplementary Material online ("RNA-seq" sheet). Gene models of nonoverlapping exonic fragments were taken from ENSEMBL 54 database. For each exonic fragment, total coverage by mapped reads in each sample was calculated with bedtools multicov tool (version 2.17.0). Total gene coverage was calculated as a sum of coverages of all nonoverlapping exonic fragments of a gene. Differential expression analysis was performed by applying default read count normalization and performing per-gene negative binomial tests, implemented in edgeR R package, with default parameters (Robinson et al. 2010). For each gene, the package provided both P values and FDRs (P values after Benjamini-Hochberg multiple testing corrections). Raw read counts and normalized read counts for each gene in each sample, as well as differential expression P values and fold-changes for MMFF, MMMF, and FFFM comparisons are provided in Supplementary table S5, Supplementary Material online.

Exploratory Statistical Analyses

Principal component analysis (PCA) was performed with a standard prcomp function in R. The genome was split into 5 kb windows, a total amount of methylation-supporting reads and total coverage was calculated for each window, thus yielding a robust estimate of DNA methylation level. Methylation values for each bins were taken as input variables, whereas each point in a multidimensional space represented a stickleback individual.

To estimate variation of DNA methylation, we calculated DNA methylation levels in 5 kb windows the same way as for PCA analysis. Standard deviation of DNA methylation levels was calculated for each genomic window within each stickleback population. Distributions of obtained standard deviations were compared between populations, and each population contained the same amount of individuals profiled.

To confirm that the observed difference in variance was not caused by different number of reads in samples, we performed read sampling. From each sample, we randomly selected the number of reads equal to the minimal number of reads among all samples and repeated read mapping, methylation calling and variance analysis as described above.

Data Availability

RRBS data was deposited to NCBI GEO under GSE82310 study accession code.

Ethics Statement

This work was approved by ethical committee of Research Center of Biotechnology RAS, Moscow, Russia

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to acknowledge Nadezhda Terekhanova for providing preprocessed allele frequencies data in sea and freshwater populations from (Terekhanova et al. 2014) and for meaningful discussion. We would like to thank Cailin Kelly for editing the manuscript. We thank Prof Konstantin Skryabin for comprehensive assistance with this project. This work was supported by Russian Science Foundation (RSF) grant #14-24-00175.

References

Barrett RDH, Paccard A, Healy TM, Bergek S, Schulte PM, Schluter D, Rogers SM. 2011. Rapid evolution of cold tolerance in stickleback. *Proc Biol Sci*. 278:233–238.

Bell MA. 2001. Lateral plate evolution in the threespine stickleback: getting nowhere fast. *Genetica* 112–113:445–461.

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302–305.

Cooper DN, Krawczak M. 1989. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet* 83:181–188.

Dmitrieva NI, Burg MB. 2007. High NaCl promotes cellular senescence. *Cell Cycle* 6:3108–3113.

Dmitrieva NI, Cui K, Kitchaev DA, Zhao K, Burg MB. 2011. DNA double-strand breaks induced by high NaCl occur predominantly in gene deserts. *Proc Natl Acad Sci U S A*. 108:20796–20801.

Dmitrieva NI, Ferraris JD, Norenburg JL, Burg MB. 2006. The saltiness of the sea breaks DNA in marine invertebrates: possible implications for animal evolution. *Cell Cycle* 5:1320–1323.

Feinberg AP, Irizarry RA. 2010. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A*. 107(Suppl 1):1757–1764.

Gao L, Geng Y, Li B, Chen J, Yang J. 2010. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant Cell Environ* 33:1820–1827.

Guo B, DeFaveri J, Sotelo G, Nair A, Merilä J. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol* 13:19.

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13:R83.

Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc Lond B Biol Sci*. 367:395–408.

James AW, Shen J, Zhang X, Asatrian G, Goyal R, Kwak JH, Jiang L, Bengs B, Culiat CT, Turner AS, et al. 2015. NELL-1 in the treatment of osteoporotic bone loss. *Nat Commun*. 6:7362.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36.

Kolka VV, Korsakova OP. 2005. Application of geological methods for dating of stone labyrinths on the White Sea coast. *Proc MSTU* 15:349–356.

Konijnendijk N, Shikano T, Daneels D, Volckaert FAM, Raeymaekers JAM. 2015. Signatures of selection in the three-spined stickleback along a small-scale brackish water - freshwater transition zone. *Ecol Evol*. 5:4174–4186.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572.

Lee S-H, Park Y, Yoon SK, Yoon J-B. 2010. Osmotic stress inhibits proteasome by p38 MAPK-dependent phosphorylation. *J Biol Chem*. 285:41280–41289.

Lescak EA, Bassham SL, Catchen J, Gelmond O, Sherbick ML, von Hippel FA, Cresko WA. 2015. Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci U S A*. 112:E7204–E7212.

Liu S, Sun K, Jiang T, Feng J. 2015. Natural epigenetic variation in bats and its role in evolution. *J Exp Biol*. 218:100–106.

Liu S, Sun K, Jiang T, Ho JP, Liu B, Feng J. 2012. Natural epigenetic variation in the female great roundleaf bat (*Hipposideros armiger*) populations. *Mol Genet Genomics* 287:643–650.

McCairns RJS, Bernatchez L. 2010. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* 64:1029–1047.

Morris MRJ, Richard R, Leder EH, Barrett RDH, Aubin-Horth N, Rogers SM. 2014. Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Mol Ecol*. 23:3226–3240.

Norman JD, Ferguson MM, Danzmann RG. 2014. An integrated transcriptomic and comparative genomic analysis of differential gene expression in Arctic charr (*Salvelinus alpinus*) following seawater exposure. *J Exp Biol*. 217:4029–4042.

Rastorguev SM, Nedoluzhko AV, Sharko FS, Boulygina ES, Sokolov AS, Gruzdeva NM, Skryabin KG, Prokhortchouk EB. 2016. Identification of novel microRNA genes in freshwater and marine ecotypes of the three-spined stickleback (Gasterosteus aculeatus). *Mol Ecol Resour*. *Available from* http://dx.doi.org/10.1111/1755-0998.12545.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Roesti M, Marius R, Walter S. 2014. Natural selection: it's a many-small world after all. *Curr Biol*. 24:R959–R962.

Smith G, Smith C, Kenny JG, Chaudhuri RR, Ritchie MG. 2015. Genome-wide DNA methylation patterns in wild samples of two morphotypes of threespine stickleback (Gasterosteus aculeatus). *Mol Biol Evol*. 32:888–895.

Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazykin GA, Kondrashov AS, Mugue NS. 2014. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet*. 10:e1004696.

Varriale A. 2014. DNA methylation, epigenetics, and evolution in vertebrates: facts and challenges. *Int J Evol Biol*. 2014:475981.

Whitmarsh AJ, Davis RJ. 2007. Role of mitogen-activated protein kinase kinase 4 in cancer. *Oncogene* 26:3172–3184.

Xia J, Han L, Zhao Z. 2012. Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. *BMC Genomics* 13(Suppl 8):S7.

Yi C, Zhang S, Liu X, Bui HTN, Hong Y. 2010. Does epigenetic polymorphism contribute to phenotypic variances in Jatropha curcas L.? *BMC Plant Biol*. 10:259.

McNew et al. BMC Evolutionary Biology (2017) 17:183
DOI 10.1186/s12862-017-1025-9

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Epigenetic variation between urban and rural populations of Darwin's finches

Sabrina M. McNew[1], Daniel Beck[2], Ingrid Sadler-Riggleman[2], Sarah A. Knutie[1], Jennifer A. H. Koop[1], Dale H. Clayton[1] and Michael K. Skinner[2*]

Abstract

Background: The molecular basis of evolutionary change is assumed to be genetic variation. However, growing evidence suggests that epigenetic mechanisms, such as DNA methylation, may also be involved in rapid adaptation to new environments. An important first step in evaluating this hypothesis is to test for the presence of epigenetic variation between natural populations living under different environmental conditions.

Results: In the current study we explored variation between populations of Darwin's finches, which comprise one of the best-studied examples of adaptive radiation. We tested for morphological, genetic, and epigenetic differences between adjacent "urban" and "rural" populations of each of two species of ground finches, *Geospiza fortis* and *G. fuliginosa*, on Santa Cruz Island in the Galápagos. Using data collected from more than 1000 birds, we found significant morphological differences between populations of *G. fortis*, but not *G. fuliginosa*. We did not find large size copy number variation (CNV) genetic differences between populations of either species. However, other genetic variants were not investigated. In contrast, we did find dramatic epigenetic differences between the urban and rural populations of both species, based on DNA methylation analysis. We explored genomic features and gene associations of the differentially DNA methylated regions (DMR), as well as their possible functional significance.

Conclusions: In summary, our study documents local population epigenetic variation within each of two species of Darwin's finches.

Keywords: Epigenetics, *Geospiza*, Copy number variation, Galápagos Islands, DNA methylation

Background

Studies of the molecular basis of evolutionary change have focused almost exclusively on genetic mechanisms. However, recent work suggests that heritable modifications to gene expression and function, independent of changes to DNA sequence, may also be involved in the evolution of phenotypes [1–3]. One of the most common of these epigenetic mechanisms is DNA methylation, i.e. the chemical attachment of methyl groups (CH_3) to nucleotides (usually a cytosine followed by a guanine- "CpG") [4]. Methylation can be induced by the environment and affect gene expression and phenotypic traits without changing the DNA sequence itself [5–8]. Importantly, some patterns of methylation are heritable, meaning they have the potential to evolve [9–14]. Indeed, because DNA methylation modifications (epimutations) are more common than genetic mutations [15], they may play a role in the rapid adaptation of individuals to new or variable environments [16].

Environmentally-induced epimutations may be a component of the adaptive radiation of closely related species to new environments [17]. For example, Skinner et al. [18] showed that epigenetic variation is significantly correlated with phylogenetic distance among five closely related species of Darwin's finches in the Galápagos Islands. Although the adaptive significance of this epigenetic variation is unknown, some of the variants are associated with genes related to beak morphology, cell signaling, and melanogensis. The results of this study suggest that epigenetic changes accumulate over macroevolutionary time and further suggest that epigenetic changes may contribute to the evolution of adaptive phenotypes.

* Correspondence: skinner@wsu.edu
[2]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA
Full list of author information is available at the end of the article

Epigenetic variation also occurs among populations within single species [15, 19–24]. Some population epigenetic studies report correlations between methylation patterns and environmental factors, suggesting that differences in methylation are involved in local adaptation to different environments [21, 22, 24]. For example, in a study of populations of two salt marsh specialist plants living along a salinity gradient, Foust et al. [24] found that ground salinity is more closely correlated with epigenetic variation than genetic variation.

The purpose of our study was to investigate epigenetic variation between populations of each of two species of Darwin's finches: the medium ground finch *(Geospiza fortis)* and the small ground finch *(G. fuliginosa)* (Fig. 1). Darwin's finches are a closely related group of about 16 species endemic to the Galápagos Islands [25–28]. Long-term studies show rapid phenotypic changes in populations of finches in response to environmental pressures, including competition [26]. The molecular basis of these phenotypic changes is poorly known. Although recent genomic studies have identified alleles in several putative genes associated with beak size and shape [28–30], most genetic markers show little differentiation among populations or species [28, 30–34].

Epigenetic variation may contribute to the phenotypic diversity of Darwin's finch populations that cannot be detected through genomic studies. As an initial test of this hypothesis, we compared components of the morphology, genetics, and epigentics in populations of finches living at El Garrapatero, a relatively undisturbed locality, to populations living near Puerto Ayora (Academy Bay), the largest town in the Galápagos Islands (Fig. 1). Hereafter, we refer to these as the "rural" and "urban" sites, respectively. The two sites, which are only 10 km apart, are both arid, lowland scrub habitat along the south and south-eastern coast of the island. Vegetative cover, based on remote sensing spectroradiometric indicies, is slightly higher at the urban site; however, cover is highly correlated between the two sites year-round (Additional file 1: Figure S1). Despite the overall ecological similarity of the sites, anthropogenic disturbance at the urban site has increased dramatically over the past fifty years [35]. Observational studies suggest urbanization has effects on finch behavior and diet: birds in the urban population incorporate novel, human foods into their diets, whereas finches in the rural popuation do not [36]. To further explore potential impacts of urbanization of Puerto Ayora on ground finches, we tested for morphological, genetic, and

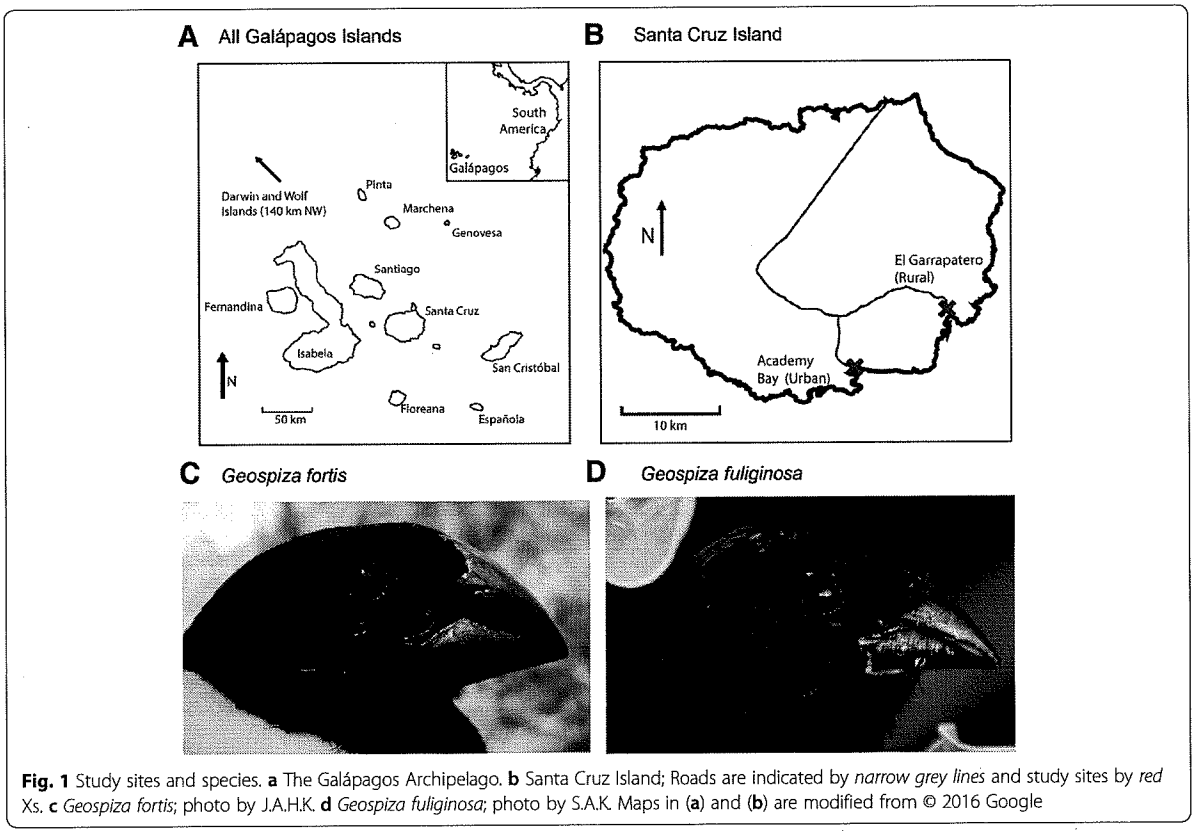

Fig. 1 Study sites and species. **a** The Galápagos Archipelago. **b** Santa Cruz Island; Roads are indicated by *narrow grey lines* and study sites by *red* Xs. **c** *Geospiza fortis*; photo by J.A.H.K. **d** *Geospiza fuliginosa*; photo by S.A.K. Maps in (**a**) and (**b**) are modified from © 2016 Google

epigenetic differences between urban and rural populations in each of two species of finches.

Methods

Study sites and species

We studied each of two populations of *G. fortis* and *G. fuliginosa* living in urban and rural environments (urban: Academy Bay; 0° 44′ 21.3″ S, 90° 18 '06.3″ W; rural: El Garrapatero; 0° 41′ 15.7″ S, 90° 13′ 18.3″ W). The two localities, which are separated by about 10 km, are both in the arid coastal zone of Santa Cruz Island (Fig. 1). *Geospiza fortis* and *G. fuliginosa* are among the most abundant species of finches at these study sites. There appears to be little movement of finches between populations. Over the course of a decade-long banding study (2002–2012), during which more than 3700 finches were captured- and more than 300 individuals recaptured- only one bird (a female *G. fortis*) was shown to have moved between the two sites (J. Raeymaekers pers. comm.).

Field work and sample collection

Finches were captured at the two study sites January–April 2008–2016. The birds were mist-netted and banded with individually numbered Monel bands in order to track individuals. They were aged and sexed using size and plumage characteristics [37]. Morphological measurements were taken from each individual including beak depth, beak width, beak length, tarsus length, wing chord, and body mass, following Grant and Grant (2014) [26], with the exception that wing chord was measured unflattened. Principle components were calculated from untransformed data for the three body measurements (mass, wing chord, and tarsus) and for the three beak measurements (length, width, and depth) to provide aggregate measures of body size and beak size and shape [38]. We evaluated morphological differences between urban and rural sites using linear mixed effects models (LMM), with site as a fixed effect, and year as a random effect to control for variation among years and investigators. Separate models were run for each morphological measurement, as well as body size (PC1 body) and beak size and shape (PC1 beak and PC2 beak). *P*-values were adjusted with a Bonferroni correction for multiple tests. Morphological analyses were run in the program RStudio using R version 3.2.1 with the packages pwr, plotrix, lme4, and lmerTest [39–42].

Blood and sperm samples for epigenetic and genetic analyses were collected from a subset of birds captured January–April 2009–2013 at the two study sites. Blood samples (<90 μl) were taken from finches via brachial venipuncture. The samples were stored on wet ice in the field and, within six hours of collection, erythrocytes were purified by centrifugation. Sperm samples (~5 μl) were taken from a subset of males. The sperm samples were obtained by gentle squeezing of the cloacal protuberance of reproductively active males. Blood erythrocytes and sperm samples were stored in a − 20 °C freezer in the Galápagos. Following each field season, they were transferred to a − 80 °C freezer in the USA for long-term storage. All field procedures were approved by the University of Utah Institutional Animal Care and Use Committee (protocols #07–08004, #10–07003 and #13–06010) and by the Galápagos National Park.

Genomic DNA preparation

Genomic DNA from finch red blood cells (erythrocytes) was prepared using the Qiagen DNAeasy Blood and Tissue Kit (Qiagen, Valencia CA). The manufacturer's instructions for nucleated blood samples were followed, but in the final DNA elution step H_2O was used instead of the buffer provided in the kit. Genomic DNA from finch sperm was prepared as follows: collected sperm suspension was adjusted to 100 μl with 1 x Phosphate Buffered Saline (PBS) then 820 μl DNA extraction buffer (50 mM Tris pH 8, 10 mM EDTA pH 8, 0.5% SDS) and 80 μl 0.1 M dithiothreitol (DTT) were added and the sample was incubated at 65 °C for 15 min. Next, 80 μl Proteinase K (20 mg/ml) were added and the sample was incubated on a rotator at 55 °C for 2 h. After incubation, 300 μl of protein precipitation solution (Promega, A795A) were added, then the sample was mixed and incubated on ice for 15 min, then spun at 4 °C at 13,000 rpm for 20 min. The supernatant was transferred to a fresh tube, then precipitated over night with the same volume of 100% isopropanol and 2 μl glycoblue at −20 °C. The sample was then centrifuged and the pellet washed with 75% ethanol, then air-dried and re-suspended in 100 μl H_2O. DNA concentration was measured using a Nanodrop Spectrophotometer (Thermo Fisher).

CNV-Seq protocol

To test for genetic differences between the urban and rural populations we sequenced DNA extracted from red blood cells (erythrocytes) and compared genetic copy number variation (CNV) [18]. CNV, defined as the changes in the number of repeat element copies of more than >1 kb of DNA, is increasingly recognized as one of the most common and functionally important markers of genetic variation [43]. The basic copy number variation (CNV) was determined through genomic sequencing of the same samples used for epigenetic analysis. Read numbers at specific loci were compared genome wide to identify CNV [18]. Erythrocyte DNA pools were generated by combining equal amounts of extracted DNA from five individuals. Each pool contained a total of 2 μg of genomic DNA. Three pools of five individuals each were created per species, per site.

Pooling samples for genomic analysis provides an accurate and cost-effective way of comparing populations [44]. Pooling decreases power, compared to sequencing individual samples. Although minor differences in copy number between populations may be missed [45], large differences between groups should be detected.

The pools were diluted to 130 μl with 1 x TE buffer and sonicated in a Covaris M220 with the manufacturer's preset program to create fragments with a peak at 300 bp. Aliquots of the pools were run on a 1.5% agarose gel to confirm fragmentation. The NEBNext DNA Library Kit for Illumina was used to create libraries for each pool, with each pool receiving a separate index primer. The libraries were sent to the University of Nevada, Reno Genomics Core for NGS on the Illumina HiSeq 2500 using a paired end PE50 application. All 6 pooled sequencing libraries for each species were run in one sequencing lane to generate approximately 30 million reads per pool. The read depth across the genome was then assessed to identify CNV and statistically assessed with a Bayesian analysis. The genome-wide paired end read depth was approximately 2× with the CNV read depth being a total of 300 to 6000 reads per CNV detected.

Methylated DNA Immunoprecipitation (MeDIP)

Following Skinner et al. [18], we used erythrocytes as a purified somatic cell type to compare differentially methylated regions (DMRs) between populations of each of the two species. For a subset of birds, we also compared DMR of germ line cells (sperm). DMRs between urban and rural populations were identified by the methylated DNA immunoprecipitation (MeDIP) of genomic DNA. MeDIP is an enrichment-based technique that uses an antibody to preferentially precipitate methylated regions of the genome that are then sequenced [46]. DMRs are identified by comparing coverage between groups of interest. MeDIP is a cost-effective way to evaluate genomic CpG methylation, and provides highly concordant results to other sequencing-based DNA methylation methods, such as bisulfite sequencing [47]. Because MeDIP surveys methylation genome-wide, it can be used to identify genomic characteristics associated with methylation. For instance, studies have found relationships between CpG density, methylation, and effects on gene transcription [6].

For analysis of erythrocytes, genomic DNA was extracted from the same individuals as used in the CNV pools. Each erythrocyte pool included five individuals and contained a total of 6 μg of genomic DNA. Sperm pools included two individuals and contained a total of 1.8 μg of genomic DNA. Three pools were generated per species, per site (for a total of $n = 6$ individuals per species, per site for sperm and $n = 15$ individuals per species per site for erythrocytes to consider biological variation of the pools and analysis). All pools were diluted to 150 μl with 1× Tris-EDTA (TE, 10 mM Tris, 1 mM EDTA) and sonicated with a probe sonicator using 5 × 20 pulses at 20% amplitude. Fragment size (200–800 bp) was verified on 1.5% agarose gel. Sonicated DNA was diluted to 400 μl with 1xTE and heated to 95 °C for 10 min, then shocked in ice water for 10 min. Next, 100 μl of 5 x immunoprecipitation (IP) buffer (50 mM Sodium Phosphate pH 7, 700 mM NaCl, 0.25% Triton X-100) and 5 μg of 5-mC monoclonal antibody (Diagenode, C15200006–500) were added and the sample was incubated on a rotator at 4 °C over night. The next day Protein A/G Agarose Beads from Santa Cruz Biotechnology, Santa Cruz CA, were prewashed with 1xPBS/0.1% BSA and re-suspended in 1 x IP buffer. Eighty μl of the bead slurry were added to each sample and incubated at 4 °C for 2 h on a rotator. The bead-DNA-antibody complex was washed 3 times with 1 x IP buffer by centrifuging at 6000 rpm for 2 min and resuspending in 1 x IP buffer. After the last wash the bead-complex was re-suspended in 250 μl of digestion buffer (50 mM Tris pH 8, 10 mM EDTA pH 8, 0.5% SDS) with 3.5 μl Proteinase K (20 mg/ml) per sample and incubated on a rotator at 55 °C for 2 h. After incubation, DNA was extracted with the same volume of Phenol-Chloroform-Isoamylcohol, then with the same volume of chloroform. To the supernatant from chloroform extraction, 2 μl glycoblue, 20 μl 5 M sodium chloride and 500 μl 100% cold ethanol were added. DNA was precipitated at −20 °C over night, then spun for 20 min at 13,000 rpm at 4 °C, washed with 75% ethanol, and air-dried. The dry pellet was re-suspended in 20 μl H_2O and concentration measured in Qubit using a Qubit ssDNA Assay Kit (Life Technologies, Carlsbad, CA).

MeDIP-Seq protocol

The next step for DMR identification involved sequencing the MeDIP DNA to identify differential methylation at specific genomic loci by assessing read numbers for the different samples. The MeDIP pools were used to create sequencing libraries for next generation sequencing (NGS) at the University of Nevada, Reno Genomics Core Laboratory using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina®, starting at step 1.4 of the manufacturer's protocol to generate double stranded DNA. After this step the manufacturer's protocol was followed. Each pool received a separate index primer. NGS was performed at the same laboratory using the Illumina HiSeq 2500 with a paired end PE50 application, with a read size of approximately 50 bp and approximately 100 million reads per pool. Two separate sequencing libraries, one rural and one urban, were run in each lane. The read depth for identified differential DNA methylated regions (DMRs) ranged from approximately 100 to >1000 total reads per DMR.

Bioinformatics

Basic read quality was verified using summaries produced by the FastQC program [48]. The reads for each sample for both CNV and DMR analyses were mapped to the zebra finch (*Taenopygia guttata*) genome using Bowtie2 [49] with default parameter options. The mapped read files were converted to sorted BAM files using SAMtools [50]. The cn.MOPS R package [51] was used to identify potential CNV. The cn.mops default information gain thresholds were used for this analysis. The cn.MOPS analysis detects CNVs by modeling read depth across all samples. The model predicts copy number for a given window based on observed read counts. The model uses a Bayesian framework to determine whether copy number for a give window differs significantly from 2. The length of the CNV is determined by comparing copy number of adjacent windows on the genome and joining those with the same copy number into one segment. A CNV call occurs when copy number for a given genomic segment varies from that of other samples. CNV detection with cn.MOPS is robust to low-coverage sequencing data (0.18–0.46 for 75 bp reads) and performs well when comparing 6 or more samples [51]. The window size used by the cn.MOPS analysis was chosen dynamically for each chromosome based on the read coverage. The chromosomes' window size ranged from approximately 25 kb to 60 kb. Only CNV that occurred in either all urban or all rural pools were compared. Although some individual pools had higher numbers of CNV, only CNV that occur red among all the pools were included in the analysis. The CNV are identified using the difference between the posterior and prior distributions from the Bayesian analysis to estimate information gain.

To identify DMR, the reference genome was broken into 100 bp windows. The MEDIPS R package [52] was used to calculate differential coverage between the urban and rural localities. The edgeR p-value [53] was used to determine the relative difference between the two localities for each genomic window. Windows with an edgeR p-value less than 10^{-3} were considered DMR. The DMR edges were extended until no genomic window with an edgeR p-value less than 0.1 remained within 1000 bp of the DMR. The DMR that included at least two windows with an edgeR p-value $<10^{-3}$ ("multiple-window DMR") were then selected for further analysis. Because no fully assembled or annotated genome exists for any Darwin's finch species, we aligned DMR with the zebra finch genome. CpG density and gene associations were then calculated for the DMR, based on alignment with the reference genome. Though we previously found high (>98%) homology between Darwin's finch and zebra finch genomes using tiling arrays [18], some differences were expected. Thus, associations of DMR with genes are likely to be under-estimates. To validate the epigenetics and gene associations, a similar analysis was also done with the draft *G. fortis* genome [54]. All the DMR sequence and genomic data obtained in the current study have been deposited in the NCBI public GEO database (GEO # GSE87825).

DMR clusters were identified with an in-house R script (www.skinner.wsu.edu under genomic data) using a 2 Mb sliding window with 50 kb intervals. DMR were annotated using the biomaRt R package [55] to access the Ensembl database [56]. The genes that overlapped with DMR were then input into the KEGG pathway search [57, 58] to identify associated pathways. A 10 kb flanking sequence was added to each DMR to consider potential localization in promoter regions of the gene as previously described [18, 59]. The DMR associated genes were manually sorted into gene classification groups by consulting information provided by the DAVID, Panther, and Uniprot databases incorporated into an internal curated database (www.skinner.wsu.edu under

Table 1 Mean (± 1SE) values for morphological characteristics of *G. fortis* and *G. fuliginosa* at rural vs. urban sites.

	G. fortis		*G. fuliginosa*	
Morphological	Rural	Urban	Rural	Urban
Character	$N = 560$	$N = 245$	$N = 171$	$N = 121$
Beak depth	11.48 ± 0.06	11.98 ± 0.09**	7.40 ± 0.04	7.42 ± 0.06
Beak width	9.89 ± 0.04	10.24 ± 0.07**	6.8 ± 0.03	6.82 ± 0.04
Beak length	11.71 ± 0.04	12.02 ± 0.07***	8.56 ± 0.04	8.46 ± 0.09
Tarsus length	21.00 ± 0.06	21.15 ± 0.09	18.83 ± 0.11	18.67 ± 0.09
Wing chord	69.3 ± 0.19	70.4 ± 0.29**	61.26 ± 0.31	61.1 ± 0.30
Body mass	21.23 ± 0.13	22.2 ± 0.23*	13.87 ± 0.15	13.76 ± 0.14
PC1 Body	−0.13 ± 0.06	0.29 ± 0.09***	0.07 ± 0.09	−0.10 ± 0.10
PC1 Beak	−0.17 ± 0.07	0.40 ± 0.11***	0.01 ± 0.09	−0.01 ± 0.15
PC2 Beak	−0.01 ± 0.02	0.02 ± 0.03	0.07 ± 0.04	−0.09 ± 0.10

Statistically significant differences between populations at $P < 0.01$, 0.001, and <0.0001 are indicated by *, ** and ***, respectively

Table 2 Differentially methylated regions (DMR) between urban and rural populations based on different cell types

Species/Cell Type	Number of Windows*					Sum of Multiple (≥2) Window DMR**
	1	2	3	4	5	
G. fortis Erythrocytes	2742	125	4	0	0	129
G. fortis Sperm	1160	97	9	3	1	110
G. fuliginosa Erythrocytes	4339	314	9	1	0	324
G. fuliginosa Sperm	1765	133	6	0	0	139

Only DMR that were significant at $P < 0.001$ are included
*DMR detected in one window alone were considered "single-window" variants (Fig. 2)
**DMR detected in two or more adjacent windows were considered "multiple-window" variants and used in subsequent analyses (Figs. 2, 3, 4, 5 and 6)

genomic data). To assess that the DMR were not false positives due to random biological variation within populations, a pairwise comparison analysis (individual pool comparison) on the genomic sequence data within the individual urban or rural sites and cell populations was performed [60].

Results

Morphology

We used 1097 birds captured between 2008 and 2016 for morphological analyses. We controlled for slight variation among years in traits by including year as a random effect in all analyses. At both sites, *G. fortis* was significantly larger than *G. fuliginosa* in all morphological traits (linear mixed-effects models $P < 0.0001$). Within species, urban *G. fortis* was significantly larger than rural *G. fortis* for all direct morphological measurements, except tarsus length (Table 1). Composite measures of *G. fortis* body and beak size (PC1 body and PC1 beak) also differed between the two sites; however, there was no difference in beak shape (PC2 beak). In contrast to *G. fortis*, *G. fuliginosa* did not differ significantly between the urban and rural populations in any of the morphological measurements or composite measures (Table 1). Because we captured more *G. fortis* than *G. fuliginosa* we did a power analysis for *G. fuliginosa*, using the effect size of morphological differences found in the *G. fortis* populations (smallest effect size = 0.256 (wing chord); largest effect size = 0.358 (beak depth)). Power for comparisons of *G. fuliginosa* appeared adequate for detecting similar effect sizes (0.69–0.91).

Copy number variation (CNV)

Mean read depth genome-wide for pools used in CNV analysis varied between 1.08× and 1.30× (overall mean = 1.22×). The total read depth of individual variants ranged from 300 to 6000. We identified unique CNV in three of six *G. fortis* pools and five of six *G. fuliginosa* pools. The total number of variants per pool ranged from 1 to 20. However, no variants were exclusive to all urban or all rural pools for either *G. fortis* or *G. fuliginosa* (Additional file 2: Fig. S2).

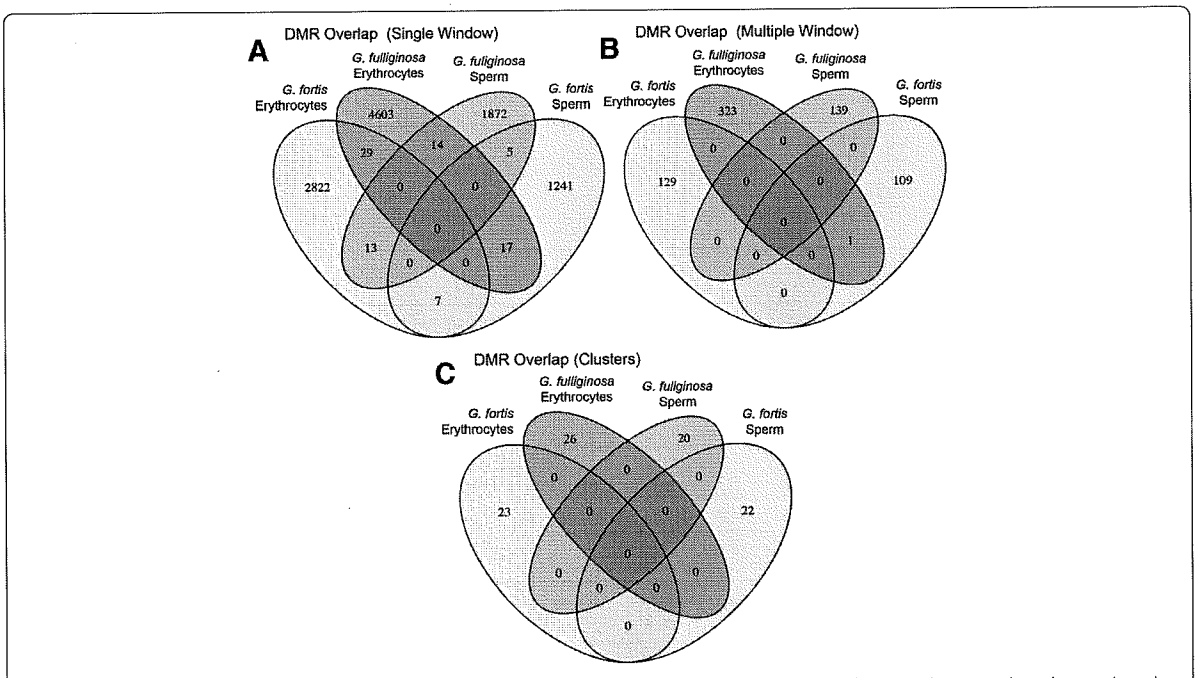

Fig. 2 DMR overlap between species and cell types. Each value is the number of differentially methylated regions between the urban and rural populations. Overlapping colors in the figure show the number of DMR that are shared between the two species or the two cell types. DMRs detected within **a** single 100 bp windows, **b** 2–5 adjacent 100 bp windows, **c** 2–7 Mb regions

Therefore, while there was variation within populations in copy number at various loci in both *G. fortis* and *G. fuliginosa* (e.g., FB2 & 12), there were no fixed differences between the urban and rural populations for either species. It is unclear why certain pools had more variants than others; however variation was consistent among chromosomes.

To control for underestimation of CNV differences due to reads that did not align to the zebra finch genome, we performed a similar analysis aligning reads to the un-assembled *Geospiza fortis* genome [54]. The average proportion of reads aligned to the *G. fortis* genome was higher (two-fold). However, we still did not find any differences in CNV between the urban and rural populations for either species of Darwin's finch. A limitation of this CNV analysis is that only large variants (>24 Kbp) can be detected reliably; smaller variants (<10 Kbp or less) may have escaped detection.

Differential DNA methylation regions (DMRs)

DMRs were found between populations for both cell types and both species (Table 2). We report the number of DMRs at *p*-value cut-offs ranging from <0.01 to <1e-05 in Additional file 3: Table S1; Additional file 4: Table S2 and Additional file 5: Table S3. The analyses reported below are restricted to DMRs significant at a level of $P < 0.001$. We evaluated differences on three "regional" scales (Fig. 2): single 100 bp window DMRs, multiple window DMRs, and "DMR clusters", i.e. statistically over-represented DMR clusters of 3–10 DMRs spanning 2–7 Mb [18] (Additional file 6: Table S4A-D). We focus on multiple-window DMRs (Additional file 4: Table S2 and Additional file 5: Table S3), i.e. DMRs detected independently in adjacent windows, because they further reduce the likelihood of false positives and provide a set of highly reproducible DMRs [18]. Multiple-window DMRs were used in the analysis of the genomic features of DMRs reported below.

There was little overlap between species or cell types in the regions that were differentially methylated between urban and rural populations (Fig. 2). A small proportion of single window DMRs (Fig. 2A) was shared between species and/or cell types. However, there were virtually no shared multiple-window DMRs (Fig. 2B) or clusters of DMRs (Fig. 2C) between species and/or cell types.

For both species and cell types, multiple-window DMRs usually were detected in only two multiple 100 bp windows; however, a limited number (<10% of total DMRs) were found in 3–5 multiple windows (Table 2). Based on extension of edges of multiple-window DMRs (extension

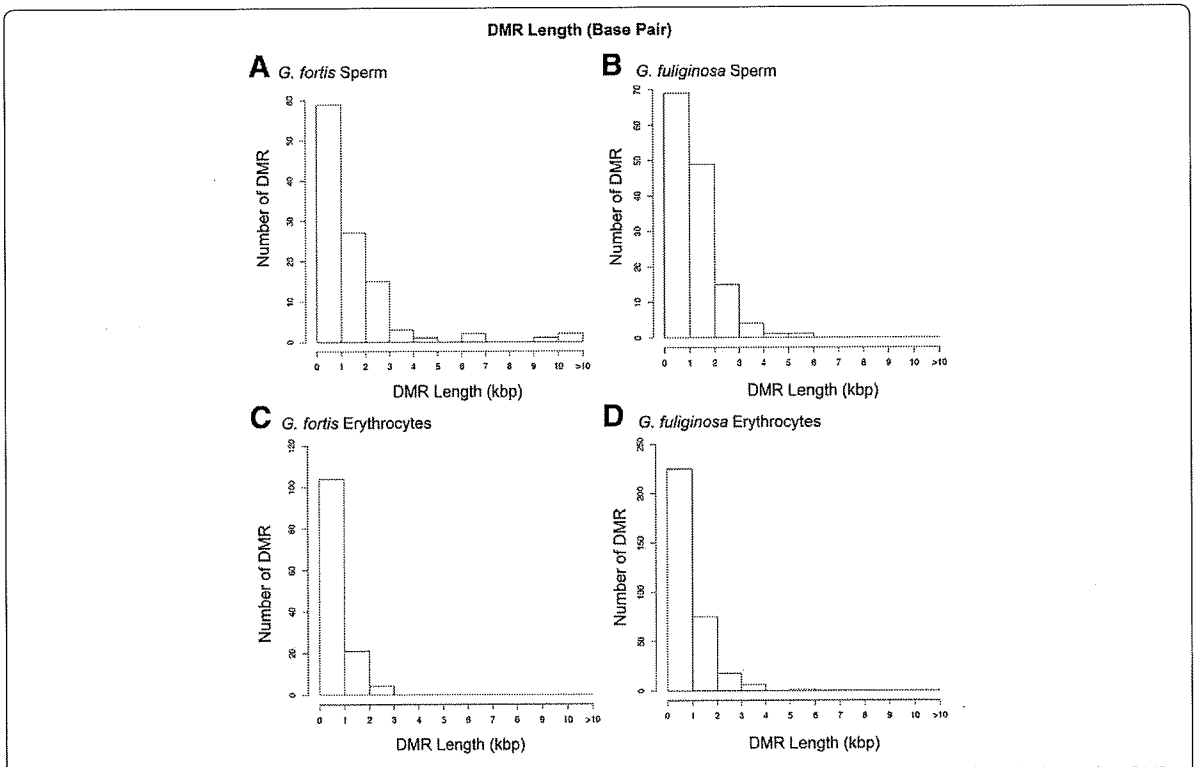

Fig. 3 DMR length (kb) in **a** *G. fortis* sperm. **b** *G. fuliginosa* sperm. **c** *G. fortis* erythrocytes. **d** *G. fuliginosa* erythrocytes. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

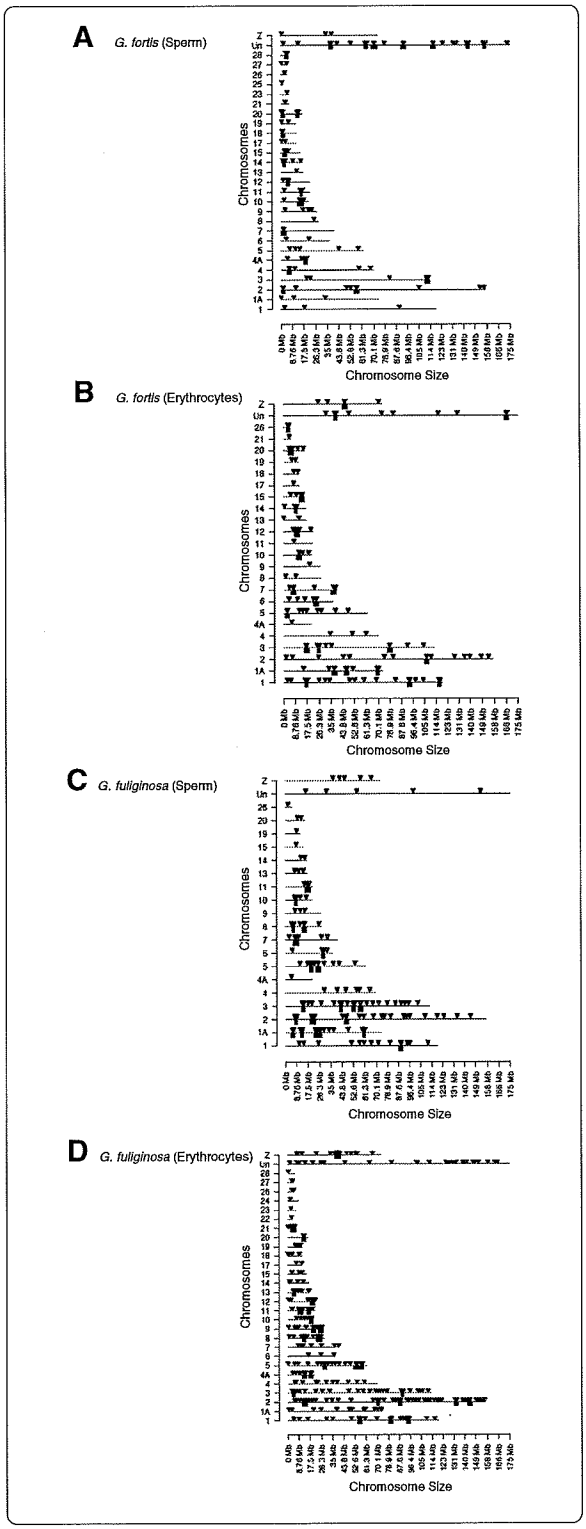

Fig. 4 Chromosomal locations of DMR identified in *Geospiza fortis* sperm **a** and erythrocytes (**b**) and *G. fuliginosa* sperm (**c**) and erythrocytes (**d**). Locations are based on alignment to the zebra finch (*Taeniopygia guttata*) genome. *Red arrowheads* indicate DMR and *black boxes* indicate DMR clusters. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are shown

of adjacent 100 bp windows with $p < 0.1$; see Methods) we estimated that most DMRs were 500–1000 bp in length (Fig. 3). Many of the DMRs in this study were clustered together, consistent with previous studies showing that DMRs are not evenly distributed across the genome [59]. Based on alignment to the zebra finch genome, we plotted the chromosomal locations of multiple-window DMRs and DMR clusters (Fig. 4). DMRs were present on all chromosomes in both sperm and erythrocytes of both species; however, the chromosomal locations of DMRs differed between the cell types and species.

We evaluated the location of DMRs with respect to nucleotide composition. CpG density was highest in DMRs of *G. fortis* sperm cells (Fig. 5A). DMRs in *G. fortis* erythrocytes and both cell types of *G. fuliginosa* were most often found in lower density CpG regions of the genome (<1 CpG site/100 bp; Fig. 5B-D). We estimated that the DMRs typically had approximately 10 CpG sites clustered within 1 kb regions.

We identified potential genes associated with DMRs through alignment with the zebra finch reference genome. DMRs within 10 kb of a gene (such that the promoter is included) have the potential to influence the gene's expression and/or pathways associated with that gene [59]. Different categories of genes were methylated in the two cell types and species (Fig. 6, specific genes listed in Additional file 7: Table S5). The most common gene categories associated with DMRs were metabolism, cell signaling and transcription (Fig. 6). Gene categories associated with DMRs differed significantly between the two species (Chi-square test, $p = 0.039$) and marginally between the two cell types (Chi-square test; $p = 0.078$). Pathway analysis (KEGG) showed DMRs associated with several genes (GALNT14, SGMS1, ENO2, PLCH2) in metabolic pathways of *G. fortis* sperm. DMRs were associated with different genes (GCLC, PRIM2, ALD1A3, AK4, ACACA) in metabolic pathways of *G. fuliginosa* sperm. *Geospiza fortis* erythrocyte DMRs were associated with genes (CACNA1H, FGF8, MRAS, RAP1A) in the MAPK signaling pathway. *Geospiza fuliginosa* erythrocyte DMRs were not associated with any particular pathway.

When the DMR data sets for both species and cell types were compared, KEGG pathways with the most DMR-associated genes were metabolic pathways, and MAPK and TGFß/BMP signaling pathways. Metabolic pathways included glycolysis, in which genes involved with pyruvate and acetate production were associated with

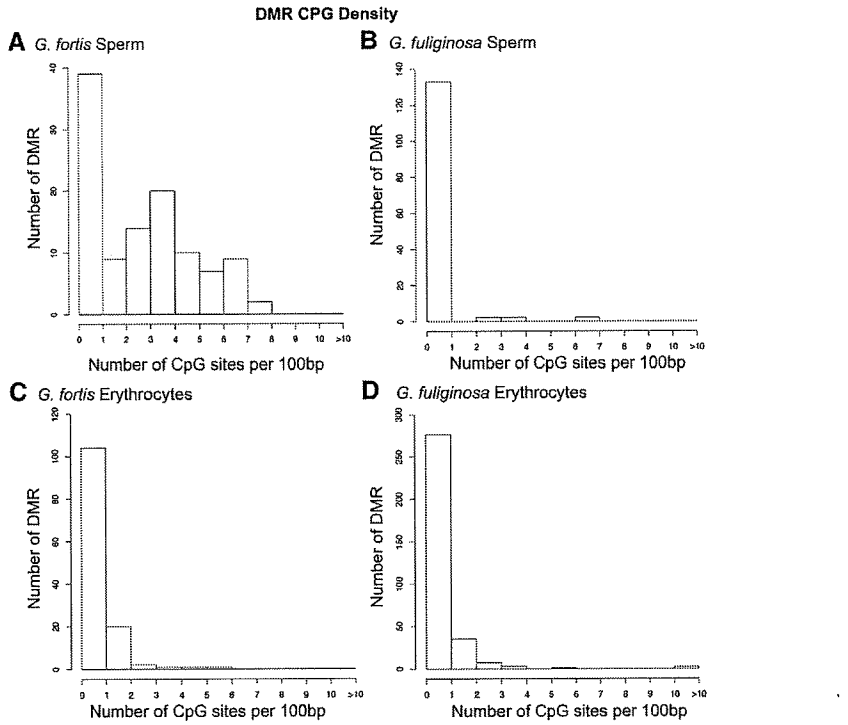

Fig. 5 The CpG density of DMR in *Geospiza fortis* sperm (**a**), *G. fuliginosa* sperm (**b**), *G. fortis* erythrocytes (**c**) and *G. fuliginosa* erythrocytes (**d**). Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

DMRs in both finch species (Additional file 8: Figure S3 and Additional file 9: Figure S4). Other metabolic pathways associated with DMRs included genes involved in purine metabolism and glycosylation (Additional file 7: Tables S5). Signaling pathways were also associated with DMRs in both species and cell types. Three genes in the TGFß/BMP pathway were associated with DMRs between *G. fuliginosa* populations (erythrocytes and sperm combined): BMP5, BMP7 and FST (Fig. 7). MAPK, a common pathway for many regulatory processes, such as cell growth, contained a high number of DMR-associated genes in both finch species (Additional file 8: Figure S3 and Additional file 9: Figure S4).

Genomic correlates of our DMR and CNV data were analyzed using the well-annotated zebra finch genome. In addition, our sequencing data were also compared to the *G. fortis* shotgun sequence database [54]. In contrast to the zebra finch genome, the *G. fortis* genome is neither assembled, nor annotated, meaning that limited data analysis is possible. The pooled individual sample read number was approximately 100 million reads for both genome analyses. The overall read alignment rate was 47–48% for the zebra finch analysis and 70–75% for the *G. fortis* genome analysis. Although previous analysis using tiling arrays suggested a 98% similarity in tiling array hybridization of the genome [18], the next generation sequencing analysis shows that more differences exist, likely in non-coding regions. The zebra finch genome analysis revealed twice the number of DMRs compared to the *G. fortis* genome analysis. This was largely due to the incomplete nature of the *G. fortis* genome. Nevertheless, analysis with both the zebra finch and *G. fortis* genomes identified epigenetic alterations between the rural and urban sites. To test whether methylation variation between sites was greater than within sites we conducted a pairwise comparison analysis (comparison of individual pools) within each species and rural or urban populations for specific cell types. We identified a number of DMRs between individual pools, which suggests that there is epigenetic variation within the study populations. However, few DMRs were found in multiple pools from the same population. Moreover, almost none of these DMRs were also found between urban and rural populations (Additional file 10: Figure S5). Thus, the DMRs identified between urban and rural populations are not an artifact of sampling within-population variation.

Discussion

Darwin's finches are well known for their phenotypic variability and evolution in response to changing environmental conditions [26]. In addition to genetic variation,

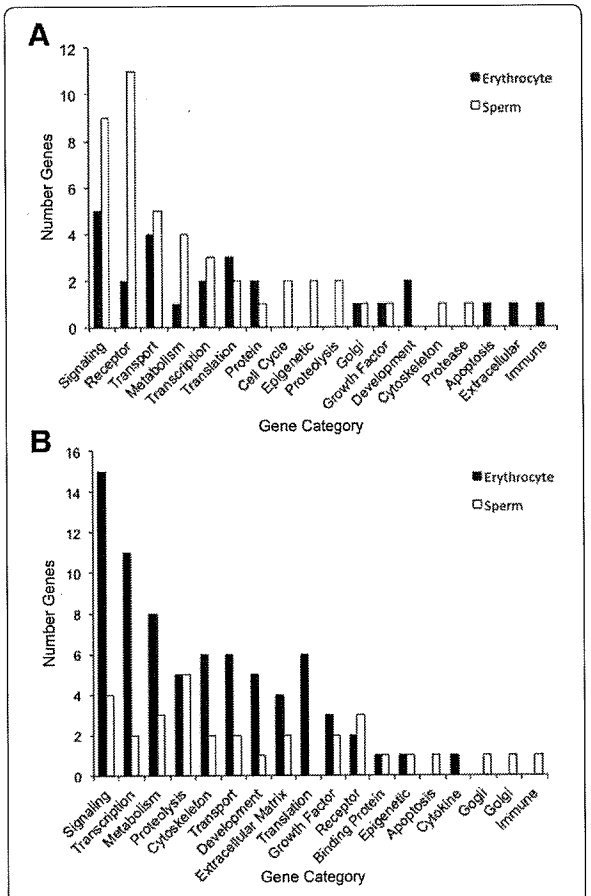

Fig. 6 Gene categories associated with DMR detected in (**a**) *G. fortis* and (**b**) *G. fuliginosa*. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

epigenetic variation - such as differential DNA methylation - may exist between natural populations living under different environmental conditions. The goal of this paper was to test for morphological, genetic, and epigenetic differences between urban and rural populations within each of two species of Darwin's finches. We found that *G. fortis* individuals at the urban site (Academy Bay) were larger than those at the rural site (El Garrapatero). In contrast, *G. fuliginosa* individuals did not differ morphologically between the sites. We did not find genetic differentiation between populations of either species based on CNV comparisons. However, we did find epigenetic (DMR) differences between urban and rural populations of both species of finches.

We found urban *G. fortis* were larger in nearly all morphological measurements compared to rural *G. fortis* (Table 1), which may be due to increased food availability at the urban site. Previous work suggests that urbanization around Academy Bay has relaxed selection on finch beak size [35, 36]. Urbanization is associated with a shift in the distribution of beak size in *G. fortis*: beak size is strongly bimodal at the rural site, whereas bimodality has decreased at the urban site concurrently with human population growth [35]. Both studies propose that increased food availability at the urban site has altered the selective landscape for *G. fortis* [35, 36]. Beak size is highly heritable in *Geospiza* finches; e.g. mid-parent vs. mid-offspring values estimate heritability of beak depth in *G. fortis* to be 0.74 [61].

In contrast, *G. fuliginosa* showed no morphological differentiation between sites (Table 1). *Geospiza fortis* is phenotypically more variable than *G. fuliginosa* on Santa Cruz Island [61]. As a result, *G. fortis* may have undergone more rapid local adaptation than *G. fuliginosa*. Although *G. fuliginosa* and *G. fortis* have overlapping dietary niches, they do show some degree of specialization [27]. It is possible that urbanization has had a greater selective effect on *G. fortis* than *G. fuliginosa*. Alternatively, morphological differences in *G. fortis* may be driven by hybridization between *G. fortis* and the slightly larger *G. magnirostris*. Hybridization between *G. fortis* and *G. magnirostris* has been documented on Santa Cruz [62]. While we have no information on relative rates of hybridization at our study sites, *G. magnirostris* is more abundant at the urban site than the rural site (4.56% of urban birds captured, compared to 1.86% of rural birds captured; unpublished data 2008–2016).

Despite differences in morphology between populations of *G. fortis*, we found no genetic differences between the urban and rural populations, based on the CNV comparisons made. Because CNV sequence coverage was limited, we may have overlooked small CNV, but larger CNV should have been detected between the two populations. CNV is a sensitive index of genetic differentiation between populations; indeed, some studies have found that CNV accounts for more genetic variation than SNPs [63–65]. Recent work has also linked CNV to rapid evolution in pepper moths [66] and primates [67].

Our study is first to explore genetic variation between populations of Darwin's finches using large–scale genomic features (CNV). Like our study, previous studies using smaller-scale genomic markers (microsatellites, nuclear introns, and mitochondrial DNA) detected little or no genetic structure within populations of either *G. fortis* or *G. fuliginosa* [31, 34, 68]. Two recent studies of genomic variation among Darwin's finches using SNPs did identify variable sites associated with variation in beak morphology [29, 30]. However, most of the genes associated with beak morphology in the two studies were different. These inconsistent results suggest that other forms of variation, such as large scale CNVs, could underlie phenotypic differences. However, our

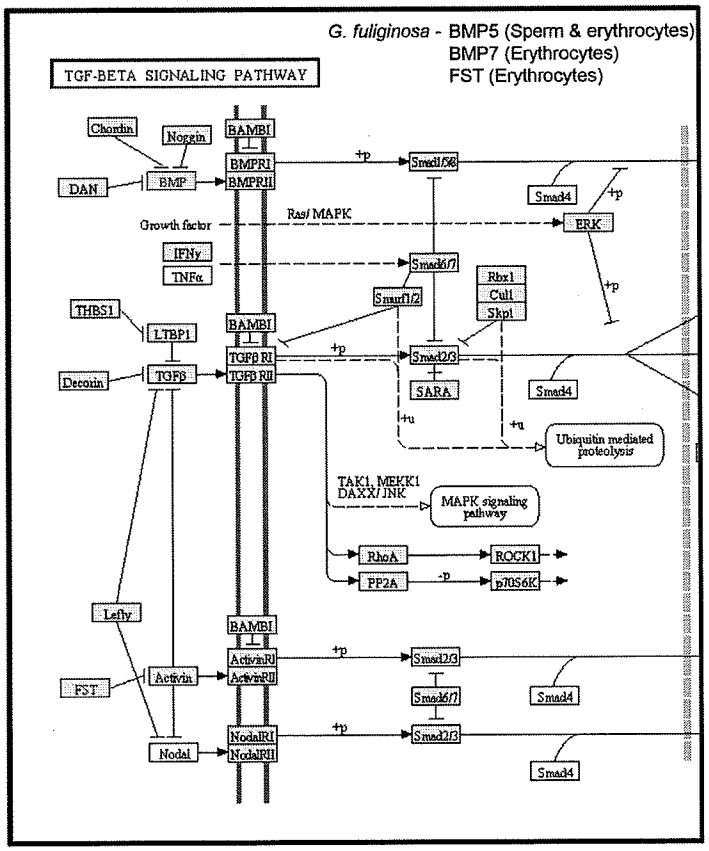

Fig. 7 TGFB/BMP pathway. Genes associated with DMR are listed and outlined in *red* in the pathway

results show that negligible large size CNV changes exist between the rural and urban populations of *G. fortis* or *G. fuliginosa*.

In contrast to our genetic results, we found a large number of epigenetic differences between urban and rural populations in both species of finches and both cell types (Fig. 2). Although DMRs were found in both species, few of the same genomic regions were differentially methylated in *G. fortis* and *G. fuliginosa*. These data suggest that methylation patterns are species-specific, even when comparing closely related species. This may mean that *G. fortis* and *G. fuliginosa* are responding to environmental changes at the urban site in different ways. The lack of overlap in DMRs between the two species may reflect differences in their diets [27]. As discussed above, dietary differences may also have contributed to the morphological differences between urban and rural populations of *G. fortis*.

Although DMRs were also found in both cell types, few of the same genomic regions were differentially methylated in sperm and erythrocytes. Because methylation is involved with cell differentiation [6, 69], some lack of similarity in erythrocyte and sperm DMR is expected. The differences between the genomic regions that were differentially methylated in sperm and erythrocytes may provide clues as to the functional significance of the epimutations. DMRs in somatic cells, such as erythrocytes, potentially reflect effects of the environment on physiology of the birds. DMRs in germ cells, such as sperm, are more likely to be transgenerationally inherited and contribute to evolution. Recent studies show that heritability of methylation variants can be high, but that this varies among loci [12]. However, without following multiple generations of individuals with known ancestry, we cannot determine which of the DMRs in our study are heritable. It is possible that many of the DMRs we detected were plastic responses to the environment. Analysis of Darwin's finches with known pedigrees - from long-term studies of banded populations - may be a way in which to distinguish heritable from non-heritable epimutations in the future.

While locations of DMRs varied between species and cell types (Fig. 4), they had genomic features in common. DMRs were typically 500–1000 bp in length (Fig. 3) and

many were clustered in 2–7 Mb regions. Most DMRs were in areas of low CpG density known as "CpG deserts" (Fig. 5). Many studies of DNA methylation have focused on the gene-silencing effects of methylation in high-density "CpG islands" near transcriptional start sites [6]. However, DMRs in other genomic regions, such as CpG deserts, can have other important effects on gene regulation and expression [6, 70]. Methylation of cytosines increases the rate of cytosine to thymine transitions [71]. Thus, over time, methylation can cause CpG-poor regions in the genome to accumulate. The persistence of conserved clusters of methylated CpG sites within CpG deserts suggests that these regions are likely conserved and under purifying selection [70]. Thus, these types of DMRs may have a functional role in regulating gene expression and could be subject to selection.

Many of the DMRs we detected were associated with metabolic and signaling genes (Fig. 6). Previous work has suggested that novel food sources at the urban site are changing the diet of finches [27]. While we did not quantify phenotypic traits related to metabolism, it is possible that DMRs associated with metabolic genes are associated with other physiological differences between the urban and rural populations.

We also found DMRs associated with genes in the bone morphogenic protein (BMP) / transferring growth factor beta (TGFß) pathway (Fig. 7). Expression of Bmp4 is related to beak shape in *Geospiza* finches [72]; however, it is unknown what factors regulate gene expression at this locus. We previously found that this pathway was differentially methylated among species of Darwin's finches [18]. These data suggest that DNA methylation may play a role in regulating expression of genes in this pathway and therefore may influence finch morphology.

Our study compared just two populations - one rural and one urban – and therefore we cannot be certain that urbanization is the key environmental change influencing finch morphology and/or epigenetics in our study. Moreover, it is possible that differences between the two populations are the result of epigenetic drift, rather than differential selection. Some dispersal of *G. fortis* between the urban and rural populations has been documented through mark-recapture studies; but it is not very common (J. Raeymaekers pers. comm.). Low levels of gene flow between populations would preclude divergence of the rural and urban populations due to drift. However, much more work is needed to understand the basis of epigenetic variation and its relationship to phenotypic variation in populations of Darwin's finches.

Conclusions

We found epigenetic differences between adjacent populations of each of two species of Darwin's finches. We do not know which of the DMRs are responses to environmental differences between the urban and rural sites, versus the result of random epigenetic drift. However, as the environmental differences between our sites are recent (<60 years) any methylation changes associated with urbanization have spread quickly. As in other recent studies [19, 20, 22], the functional relationship between environmental and epigenetic variation is not well understood. Nevertheless, these results are consistent with a potential role of epigenetic variation in rapid adaptation to changing environments. Future studies are needed to determine what effects DMR have on phenotypes, and to what extent these methylation patterns may play a role in evolution.

Additional files

Additional file 1: Figure S1. Comparison of vegetative cover at the rural site (El Garrapatero) versus urban site (Puerto Ayora, Academy Bay) over the course of the study. Cover was dervied from Normalized Difference Vegetative Index (NDVI) values generated from satellite imagery (ORNL DAAC. 2008. MODIS Collection 5 Land Products Global Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed May 08, 2017 http://dx.doi.org/10.3334/ORNLDAAC/1241). Values range from 0-1 with 1 reprensenting the highest vegetation cover. (PDF 850 kb)

Additional file 2: Figure S2. Copy number variation (CNV) between the rural and urban populations. (A) CNV analysis summary for the *G. fortis* erythrocytes showing read depth and alignment, and CNV numbers per pool with chromosomes containing CNV indicated, and no overlap between rural and urban pools indicated. (B) CNV analysis summary for the *G. fuliginosa* erythrocytes with Read Mapping Summary, overall CNV per pool and chromosome, and no overlapping CNV identified. (PDF 20 kb)

Additional file 3: Table S1. The number of DMR detected at single window and multiple window scales at increasing levels of significance. (PDF 61 kb)

Additional file 4: Table S2. Description of multiple-window DMR detected in *G. fortis* sperm **(A)** and erythrocytes **(B)**. Description includes DMR name, chromosome number, DMR start site, length in base pair (bp), number of multiple sites, minimum p-value, CpG number per sequence length, CpG density (CpG number / 100 bp) and DMR gene association. "NA" indicates DMR associated with a gene that did not align to the zebra finch reference genome. (PDF 126 kb)

Additional file 5: Table S3. Description of multiple-window DMR detected in *G. fuliginosa* sperm **(A)** and erythrocytes **(B)**. Description includes DMR name, chromosome number, DMR start site, length in base pair (bp), number of multiple sites, minimum p-value, CpG number per sequence length, CpG density (CpG number / 100 bp) and DMR gene association. "NA" indicates DMR associated with a gene that did not align to the zebra finch reference genome. (PDF 154 kb)

Additional file 6: Table S4. Description of DMR clusters detected in *G. fortis* sperm **(A)** and erythrocytes **(B)** and *G. fuliginosa* sperm **(C)** and erythrocytes **(D)**. Description includes DMR in cluster, chromosome number, cluster start site, cluster stop site, length in bp, and minimum *p*-value. (PDF 103 kb)

Additional file 7: Table S5. Gene associations with DMR detected in *G. fortis* sperm **(A)** and erythrocytes **(B)** and *G. fuliginosa* sperm **(C)** and erythrocytes **(D)**. Description includes DMR name, gene symbol, entrez gene identification, chromosome number, start position site, ensemble gene identification number, gene description and gene classification category. (PDF 225 kb)

Additional file 8: Figure S3. MAPK signaling pathway. Genes associated with DMR are listed and outlined in red in the pathway. (PDF 109 kb)

Additional file 9: Figure S4. Glycolysis metabolism pathway. Genes associated with DMR are listed and outlined in red in the pathway. (PDF 66 kb)

Additional file 10: Figure S5. DMRs identified in pairwise comparison of pools within populations: **(A)** *G. Fuliginosa* RBC urban analysis, **(B)** *G. fuliginosa*-RBC rural analysis, **(C)** *G. fortis* RBC urban analysis, and **(D)** *G. fortis* rural analysis. Numbers indicate DMRs between urban (U) or rural (R) individual pools (1-3). "Full analysis" are DMRs identified between urban and rural pools. DMRs identified in the full analysis were found independently of within-site variation. (PDF 98 kb)

Abbreviations
BMP: bone morphogenic protein; CNV: copy number variation; DDT: dithiothreitol; DMR: differentially DNA methylated region; IP: immunoprecipitation; LMM: linear mixed effects models; NGS: next generation sequencing; PBS: Phosphate Buffered Saline; TGFß: transferring growth factor beta

Acknowledgments
We acknowledge the advice and critical review of Dr. Eric Nilsson (WSU). The following people contributed to sample collection: Céline Le Bohec, Sarah Bush, Roger Clayton, Miriam Clayton, Oliver Tiselma, Elena Arriero, Andrew Bartlow, Daniela Vargas, Emily DiBlasi, Jordan Herman, Pricilia Espina, Kiyoko Gotanda, Sofia Carvajal, Joost Raeymakers and Janaí Yepez. We thank Ms. Jayleana Barton for molecular technical assistance and Ms. Heather Johnson for assistance in preparation of the manuscript. The research was supported by a Templeton grant to MKS, National Science Foundation grants DEB-0816877 and DEB-1342600 to DHC, and an NSF Graduate Research Fellowship to SMM. Dr. Sarah A. Knutie's present address: Department of Ecology and Evolutionary Biology, Storrs, CT 06269-3043, USA. Dr. Jennifer A. H. Koop's present address: Biology Department, University of Massachusetts-Dartmouth, Dartmouth, MA 02747-2300, USA.

Funding
The research was supported by a Templeton grant to MKS, a National Science Foundation grant to DHC, and an NSF Graduate Research Fellowship to SMM.

Availability of data and materials
All the DMR sequence and genomic data obtained in the current study have been deposited in the NCBI public GEO database (GEO # GSE87825).

Author contributions
DHC and MKS designed the study; SMM, SAK and JAHK collected the samples, DB and ISR analyzed the genomic data, SMM, DHC and MKS analyzed the data and wrote the manuscript with help from the other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate
All field procedures were approved by the University of Utah Institutional Animal Care and Use Committee (protocols #07–08004, #10–07003 and #13–06010) and by the Galápagos National Park.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Biology, University of Utah, Salt Lake City, UT 84112-0840, USA. [2]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA.

Received: 26 January 2017 Accepted: 26 July 2017
Published online: 24 August 2017

References
1. Bossdorf O, Richards CL, Pigliucci M. Epigenetics for ecologists. Ecol Lett. 2008;11(2):106–15.
2. Day T, Bonduriansky R. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. Am Nat. 2011;178(2):E18–36.
3. Robertson M, Richards C. Non-genetic inheritance in evolutionary theory - the importance of plant studies. Non-Genetic Inherit. 2015;2:3–11.
4. Angers B, Castonguay E, Massicotte R. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol Ecol. 2010;19(7):1283–95.
5. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003; 33(Suppl):245–54.
6. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.
7. Duncan EJ, Gluckman PD, Dearden PK. Epigenetics, plasticity, and evolution: how do we link epigenetic change to phenotype? J Exp Zool B Mol Dev Evol. 2014;322(4):208–20.
8. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet. 2007;8(4):253–62.
9. Crews D, Gore AC, Hsu TS, Dangleben NL, Spinetta M, Schallert T, Anway MD, Skinner MK. Transgenerational epigenetic imprints on mate preference. Proc Natl Acad Sci U S A. 2007;104(14):5942–6.
10. Richards CL, Bossdorf O, Pigliucci M. What role does heritable epigenetic variation play in phenotypic evolution? Bioscience. 2010;60:232–7.
11. Latzel V, Zhang Y, Karlsson Moritz K, Fischer M, Bossdorf O. Epigenetic variation in plant responses to defence hormones. Ann Bot. 2012;110(7):1423–8.
12. Janowitz Koch I, Clark MM, Thompson MJ, Deere-Machemer KA, Wang J, Duarte L, Gnanadesikan GE, McCoy EL, Rubbi L, Stahler DR, et al. The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. Mol Ecol. 2016;25(8):1838–55.
13. Verhoeven KJ, vonHoldt BM, Sork VL. Epigenetics in ecology and evolution: what we know and what we need to know. Mol Ecol. 2016;25(8):1631–8.
14. Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. Trends Endocrinol Metab. 2010;21(4):214–22.
15. Richards CL, Schrey AW, Pigliucci M. Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. Ecol Lett. 2012;15(9):1016–25.
16. Liu QA. The impact of climate change on plant epigenomes. Trends Genet. 2013;29(9):503–5.
17. Flatscher R, Frajman B, Schonswetter P, Paun O. Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? Genet Res Int. 2012;2012:698421.
18. Skinner MK, Guerrero-Bosagna C, Haque MM, Nilsson EE, Koop JAH, Knutie SA, Clayton DH. Epigenetics and the evolution of Darwin's Finches. Genome Biology & Evolution. 2014;6(8):1972–89.
19. Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. Epigenetic variation in mangrove plants occurring in contrasting natural environment. PLoS One. 2010;5(4):e10326.
20. Liu S, Sun K, Jiang T, Ho JP, Liu B, Feng J. Natural epigenetic variation in the female great roundleaf bat (Hipposideros Armiger) populations. Mol Gen Genomics. 2012;287(8):643–50.
21. Gugger PF, Fitz-Gibbon S, PellEgrini M, Sork VL. Species-wide patterns of DNA methylation variation in Quercus Lobata and their association with climate gradients. Mol Ecol. 2016;25(8):1665–80.
22. Lea AJ, Altmann J, Alberts SC, Tung J. Resource base influences genome-wide DNA methylation levels in wild baboons (Papio Cynocephalus). Mol Ecol. 2016;25(8):1681–96.
23. Zhao Y, Tang JW, Yang Z, Cao YB, Ren JL, Ben-Abu Y, Li K, Chen XQ, Du JZ, Nevo E. Adaptive methylation regulation of p53 pathway in sympatric speciation of blind mole rats, Spalax. Proc Natl Acad Sci U S A. 2016;113(8):2146–51.
24. Foust CM, Preite V, Schrey AW, Alvarez M, Robertson MH, Verhoeven KJ, Richards CL. Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. Mol Ecol. 2016;25(8):1639–52.

25. Podos J. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature. 2001;409(6817):185–8.
26. Grant PR, Grant BR. 40 years of evolution. Princeton, NJ: Princeton University Press; 2014.
27. De Leon LF, Podos J, Gardezi T, Herrel A, Hendry AP. Darwin's finches and their diet niches: the sympatric coexistence of imperfect generalists. J Evol Biol. 2014;27(6):1093–104.
28. Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerova M, Rubin CJ, Wang C, Zamani N, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518(7539):371–5.
29. Lamichhaney S, Han F, Berglund J, Wang C, Almen MS, Webster MT, Grant BR, Grant PR, Andersson L. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science. 2016;352(6284):470–4.
30. Chaves JA, Cooper EA, Hendry AP, Podos J, De Leon LF, Raeymaekers JA, MacMillan WO, Uy JA. Genomic variation at the tips of the adaptive radiation of Darwin's finches. Mol Ecol. 2016;25(21):5282–95.
31. de Leon LF, Bermingham E, Podos J, Hendry AP. Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. Philos Trans R Soc Lond Ser B Biol Sci. 2010;365(1543):1041–52.
32. Sato A, O'HUigin C, Figueroa F, Grant PR, Grant BR, Tichy H, Klein J. Phylogeny of Darwin's finches as revealed by mtDNA sequences. Proc Natl Acad Sci U S A. 1999;96(9):5101–6.
33. Petren K, Grant B, Grant P. A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proc R Soc Lond B. 1999;266:321–9.
34. Farrington HL, Lawson LP, Clark CM, Petren K. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. Evolution. 2014;68(10):2932–44.
35. Hendry AP, Grant PR, Rosemary Grant B, Ford HA, Brewer MJ, Podos J. Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. Proc Biol Sciences / R Soc. 2006;273(1596):1887–94.
36. De Leon LF, Raeymaekers JA, Bermingham E, Podos J, Herrel A, Hendry AP. Exploring possible human influences on the evolution of Darwin's finches. Evolution. 2011;65(8):2258–72.
37. Grant P. Ecology and evolution of Darwin's finches. Princeton, NJ: Princeton University Press; 1986.
38. Grant PR, Grant BR. Evolution of character displacement in Darwin's finches. Science. 2006;313(5784):224–6.
39. Kuznetsova A, Brockhoff P, Christensen R: lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0–33. 2016.
40. Bates D, Maechler M, Bolker B, Walker S. lme4: linear mixed-effects models using Eigen and S4. J Stat Softw. 2015;67:1–48.
41. Lemon J. Plotrix: a package in the red light district of R. R-News. 2006;6:8–12.
42. Champely S: pwr: Basic functions for power analysis. . R package version 12–1 2017.
43. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, et al. Copy number variation: new insights in genome diversity. Genome Res. 2006;16(8):949–61.
44. Schlotterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15(11):749–63.
45. Zhang W, Carriquiry A, Nettleton D, Dekkers JC. Pooling mRNA in microarray experiments and its effect on power. Bioinformatics. 2007;23(10):1217–24.
46. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc. 2012;7(4):617–36.
47. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010;28(10):1097–105.
48. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–9.
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
51. Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, Hochreiter S. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. 2012;40(9):e69.
52. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014;30(2):284–6.
53. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
54. Li B, Li H, Parker P, Wang J. The genome of Darwin's finch (Geospiza Fortis). GigaDB. 2012;
55. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91.
56. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. Ensembl 2015. Nucleic Acids Res. 2015;43(Database issue):D662–9.
57. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
58. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199–205.
59. Skinner MK, Manikkam M, Haque MM, Zhang B, Savenkova M. Epigenetic Transgenerational inheritance of somatic Transcriptomes and epigenetic control regions. Genome Biol. 2012;13(10):R91.
60. Shnorhavorian M, Schwartz SM, Stansfeld B, Sadler-Riggleman I, Beck D, Skinner MK. Differential DNA Methylation Regions in Adult Human Sperm Following Adolescent Chemotherapy: Potential for Epigenetic Inheritance. PloS One. 2017;12(2):e0170085.
61. Grant P, Grant R. How and why species multiply: the radiation of Darwin's finches Princeton. NJ: Princeton University Press; 2008.
62. Huber SK, De Leon LF, Hendry AP, Bermingham E, Podos J. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proceedings Biological sciences / The Royal Society. 2007;274(1619):1709–14.
63. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444–54.
64. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007;8(8):639–46.
65. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008;40(10):1166–74.
66. Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102–5.
67. Niu AL, Wang YQ, Zhang H, Liao CH, Wang JK, Zhang R, Che J, Su B. Rapid evolution and copy number variation of primate RHOXF2, an X-linked homeobox gene involved in male reproduction and possibly brain function. BMC Evol Biol. 2011;11:298.
68. Petren K, Grant PR, Grant BR, Keller LF. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Mol Ecol. 2005;14(10):2943–57.
69. Bestor TH. The DNA methyltransferases of mammals. Hum Mol Genet. 2000; 9(16):2395–402.
70. Skinner MK, Guerrero-Bosagna C. Role of CpG deserts in the epigenetic Transgenerational inheritance of differential DNA methylation regions. BMC Genomics. 2014;15(1):692.
71. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. Hum Genet. 1988;78(2):151–5.
72. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. Bmp4 and morphological variation of beaks in Darwin's finches. Science. 2004;305(5689):1462–5.