EVOLUTION & DEVELOPMENT 17:4, 240–247 (2015)
DOI: 10.1111/ede.12128

Gene-based polymorphisms reveal limited genomic divergence in a species with a heritable life-history dimorphism

Christina Zakas* and Matthew V. Rockman

New York University, New York, NY, USA

*Authors for correspondence (e-mail: cz12@nyu.edu and mrockman@nyu.edu)

SUMMARY Understanding the genetic basis of life-history traits is a long-standing goal of evolutionary biology. Many closely related species have contrasting life-history strategies, suggesting that the switches in early development that lead to divergent life-histories evolve quickly and frequently. Life-history changes that originate in early development have profound downstream effects on a species' morphology, ecology, genetic diversity, and even speciation rate. How do such transitions in development mode occur, and what is the underlying genetic architecture? To begin to address these questions, we investigated genetic variation in an emerging model in developmental evolution, the polychaete *Streblospio benedicti,* which has two contrasting and highly heritable offspring types. We compare transcript-based SNP genotypes of individuals of the two development modes to determine the extent of genomic differentiation between them. We find that there is extensive allele sharing across the two types, and minimal fixed differences. We use the site frequency spectrum to fit demographic models to our data and determine that there is recent gene flow between developmental morphs. Our data suggest that the evolution of a genetic developmental dimorphism is not associated with longstanding genetic isolation or genomically extensive divergence. Rather, differences at developmentally important loci, or modest allele-frequency differences at many loci, may be responsible for the drastic life-history differences.

INTRODUCTION

Transitions in life-history mode are a ubiquitous feature in animal evolution. Sister species that differ in developmental mode are common (Raff 1996), and understanding how these developmental trajectories evolve remains a major goal for evolutionary biology. Ideally, studies of developmental divergence would focus on transitions in progress; analysis of intraspecific variation eliminates confounding effects of species differences. Yet instances of true developmental dimorphism are rarely found in nature. The marine polychaete *Streblospio benedicti* is an exceptional case, with a morphologically dramatic and highly heritable developmental dimorphism (Levin 1984; Levin et al. 1991). *S. benedicti* is a particularly attractive system for investigating developmental evolution on the genomic level because females differ dichotomously in the size of eggs they produce and the developmental and morphological features of the resulting larvae. This rare instance of developmental dimorphism—termed *poecilogony*—presents the chance to investigate early development in a population–genetic context. Here, we use *S. benedicti* to estimate the extent of genome-wide divergence associated with variation in developmental mode. Understanding the genetic changes that underlie poecilogony can provide important insights into the rules that govern major developmental transitions across animal phylogeny, which in the age of next generation sequencing technology makes *S. benedicti* a powerful new model for developmental evolution.

The underlying genetic mechanisms that maintain dimorphism in offspring type are unknown. Theory suggests that a genetic architecture involving few loci of large effect, possibly linked in a chromosomal inversion, best suits the maintenance of divergent phenotypes in the face of gene flow (Charlesworth et al. 1997; Feder and Nosil 2010; Yeaman and Whitlock 2011). Recent experimental data, however, show that polygenic systems may also underlie phenotypic dimorphism (Arnegard et al. 2014). Because *S. benedicti* produces morphologically distinct offspring types as a result of heritable variation, it provides a powerful test of alternative genetic architectures in a system that couples differences in morphology, fecundity, and dispersal.

S. benedicti is a common spionid polychaete that inhabits muddy marine and estuarine sediments. Adults burrow in tubes they create from the sediment, and long-distance dispersal is primarily, if not exclusively, achieved in the larval phase. Females reproduce in one of two distinct patterns: females produce clutches of tens of large eggs, each ~200 μm diameter, giving rise to lecithotrophic (non-obligately feeding) larvae, or clutches of hundreds of small eggs, each ~100 μm diameter, giving rise to planktotrophic (obligately feeding) larvae (Zakas and Rockman 2014; Fig. 1). Egg size is fixed for each female and does not change with maternal condition or environment (Levin

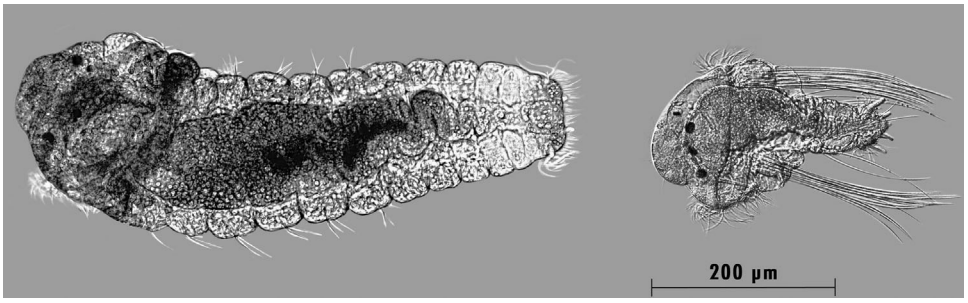

Fig. 1. A representative lecithotrophic larva from Long Beach CA (left) and a planktotrophic larva from Bayonne NJ (right).

1984; Bridges 1993; Levin and Bridges 1994). Embryos of both types develop into larvae in maternal brood pouches, after which planktotrophic larvae spend a week or longer feeding in the plankton before settling. Lecithotrophic larvae can settle immediately at release. Planktotrophic larvae develop characteristic morphological features such as larval chaetae, which likely function in defense, and pronounced anal cirri (Fig. 1). The much larger lecithotrophic larvae lack chaetae and have a heterochronic delay in timing of gut and coelom development (Pernet and McHugh 2010; Gibson et al. 2010). Differences between larval types in relative blastomere sizes are already evident in early cleavage-stage embryos (McCain 2008).

Developmental mode is generally uniform within local populations (Zakas and Wares 2012), making population and larval mode confounded. Both larval types are found in their native range on the East Coast of North America, and there is no clear latitudinal or environmental correlation with larval type. On the West Coast, only lecithotrophic worms are found, likely a result of very recent ($\sim$100 ya) introductions (Carlton 1979; Zakas and Wares 2012). Comparisons across East and West coast populations will therefore reflect patterns of long-term historical gene flow and only very recent isolation.

While early studies have identified the species as a true case of poecilogony, with more than one mode of development within a species (Levin and Creed 1986; Levin and Huggett 1990; Levin et al. 1991; Levin and Bridges 1994), the underlying genetic mechanisms that maintain this offspring dimorphism are not known. Among the possible genetic architectures are extensive fixed differences, perhaps involving large genomic inversions; a small number of fixed differences in a one or a few key regulatory genes; and an absence of relevant fixed differences, with phenotypic differentiation dependent on allele frequency differences at many loci. Previous population genetic studies of *S. benedicti* show that geographic structure between local populations is greater than expected for a species with high dispersal potential, but genetic structure between individuals of different larval modes, by contrast, is small (Zakas and Wares 2012). That work developed SNPs from only one population, lecithotrophic animals from Long Beach, CA, and therefore could not identify SNPs fixed for different alleles between larval types. Only three loci—the mitochondrial genome and two nuclear loci—have been examined without this form of ascertainment bias. Mitochondrial haplotypes are shared across larval types (Schulze et al. 2000), and of two nuclear loci compared between populations of planktotrophic worms from Bayonne, NJ, and lecithotrophic worms from Long Beach, CA, genetic differentiation was absent at one locus and slight at the other (Rockman 2012). Further crosses between adults from these two populations demonstrate that differences in one aspect of larval chaetae are consistent with divergence at a modest number of loci, as opposed to an accumulation of differentiation throughout the genome (Zakas and Rockman 2014). These studies all suggest that there is little genomic divergence between individuals of the two larval modes, an interesting finding for a geographically-associated dimorphism. Yet, the extent of genome-wide divergence has not been characterized at the molecular level.

Here, as a preliminary step toward dissecting the genetic basis of poecilogony, we characterize the population-level divergence between adults of the two developmental types. While we ultimately aim to uncover the specific gene regions that form the basis for poecilogony, here we focus on estimating the extent of divergence that accompanies this dimorphism across populations. For practical reasons, we limit our analysis to a subset of the genome, the regions transcribed in adult heads. We selected this tissue with the expectation that gene expression in adult head should be similar between animals of the two developmental morphs, allowing us to use transcriptome sequencing to achieve reduced representation coverage of the same genomic fraction. Our intraspecific analysis of populations with contrasting developmental modes can provide insights into the genetic changes that influence developmental evolution by demonstrating the extent of genomic divergence between populations with alternative developmental modes.

METHODS

Worms were collected from two representative and readily accessible populations: four planktotrophic worms were

sampled from Bayonne, NJ (Newark Bay), and four lecithotrophic worms were from the introduced population at Long Beach, CA. These populations have been sampled repeatedly in past years and larval mode has remained consistent, indicating that there has not been recent population turnover in development type. Due to its recent establishment, Long Beach is a consistent and exclusively lecithotrophic population that has not experienced recent gene flow from planktotrophic populations. We isolated RNA from whole heads, amputated anterior to the collar. We used the NuGen Ovation RNA-Seq System V2 to prepare cDNA from total RNA and the NuGen Encore NGS Multiplex System 1 to prepare barcoded mRNA libraries. These were sequenced using Illumina PE-100 chemistry. We assembled raw reads from six of the eight total worms to make a consensus alignment using *Trinity* (Grabherr et al. 2011). Two individuals (one Bayonne female, one Long Beach male) were excluded from the consensus assembly because they either had few sequences or generally low mapping quality to the consensus.

We verified our assembly quality by ensuring we can recover core genes with known conservation across eukaryotes using CEGMA (Parra et al. 2007). In addition, we ran *Pfam*(*A/B*) (Finn et al. 2008), *InterProScan* (Zdobnov and Apweiler 2001), and *blastx* (Altschul et al. 1990) to determine the proportion of components that have known functional annotations. We used the program *Kraken* (Wood and Salzberg 2014) to determine bacterial, archaeal and viral contaminants based on the reference dataset from NCBI's RefSeq.

We used *GATK* (McKenna et al. 2010) to map reads from all eight individuals to the consensus chrysalis assembly generated in *Trinity*, and we filtered on typical quality criteria for SNPs (Filter mapping quality <40, Fisher Strand value >40, ReadsPosRankSum <-8.0, Quality by depth <2). These SNPs were further filtered using VCF tools (Danecek et al. 2011) with a higher stringency to recover only high confidence SNPs (read depth = 10–500, minimum mean depth averaged across samples = 10, minimum base quality = 30, minimum genotype quality = 20). We used only biallelic SNPs, and we excluded SNPs at which every sample was called heterozygous, as this is likely due to duplication or assembly bias.

We calculated the number of non-singleton SNPs, meaning the minor allele occurs more than once. These non-singleton SNPs were then sorted into three mutually exclusive categories: fixed differences, private SNPs that segregate only in one population, and shared polymorphisms that segregate within each of the two populations. To determine if our results are consistent with a null hypothesis of a single panmictic, randomly mating population, we performed an exact permutation test. We permuted the eight individuals into all possible combinations of four and four (35 possible combinations) and recalculated the number of SNPs in each category. As a test statistic, we used the log likelihood of the data configuration under the null hypothesis. For non-singleton SNP *i,* at which $N_{B,i}$ Bayonne individuals ($2N_{B,i}$ chromosomes) and $N_{L,i}$ Long Beach individuals have been observed, the minor allele is observed k_i times, and n_i of those observations are among Bayonne individuals, then the probability that x_i of the Bayonne alleles are of the minor allele state follows from the hypergeometric distribution, and the log likelihood of the entire non-singleton dataset is given by

$$\ln L = \sum_{i}^{SNPs} \ln \frac{\binom{k_i}{x_i}\binom{2N_{B,i}+2N_{L,i}-k_i}{2N_{B,i}-x_i}}{\binom{2N_{B,i}+2N_{L,i}}{2N_{B,i}}}$$

We also determined the distribution of ln$L$ under the null hypotheses of panmixia and linkage equilibrium by simulating SNP datasets with the observed allele frequency spectrum. The permutations, simulations, and calculations were performed in R (R Core Development Team 2011).

We modeled the long-term demographic history of the two populations from the site frequency spectrum (SFS) of our high-quality SNPs, those that were genotyped successfully in all 8 individuals. We use a diffusion approximation approach, implemented in ∂a∂i v1.6.3 (Gutenkunst et al. 2009). Because our high-quality SNP set contained genotypes for all individuals, there was no need to project down to a smaller sample size. Without ancestral-state information, we used a folded SFS. Input parameters were estimated as suggested, from trials of model-fitting using simple models and a wide range of starting values (upper and lower parameter bounds are 0.001–100 for all parameters). The IM (isolation-with-migration) demographic model was a better fit than the standard neutral model (where populations do not diverge) and was used to fit the data set and estimate divergence parameters for the two populations. Start values were randomized and grid size was selected to be slightly larger than the sample size. We used likelihood ratio testing to pick the best model parameters that fit the data (Table 1). Linkage disequilibrium between loci decays rapidly in *S. benedicti* (Rockman 2012), but many SNPs in our dataset occur within the same transcriptome assembly component, violating the assumption of linkage equilibrium. To minimize known linkage effects in our data, we generated 100 random samples of the data including only one SNP per assembly component. We then fit model parameters to these sampled datasets to generate confidence intervals.

RESULTS

We recovered ~4 million reads per individual (except the one Bayonne female excluded from the consensus assembly, which had only 800,000 reads). The alignment resulted in a 87MB assembly (using the longest isoforms only as a the representative

Table 1. Annotation for 19 components containing the 31 fixed SNPs observed at sites with all 8 worms genotyped

Component	Component length (bp)	High quality SNPs	Fixed high quality SNPs	NCBI reference	E value	Top hit description; top hit organism
s_352610	35034	92	1	XM_005105626.1	3 e−109	PREDICTED: radial spoke head protein 4 homolog A-like (LOC101857895), transcript variant X2, mRNA; *Aplysia californica*
s_355974	22210	74	2	XM_004523623.1	0	PREDICTED: serine/threonine-protein kinase minibrain-like (LOC101454974), transcript variant X3, mRNA; *Ceratitis capitata*
s_354557	7959	37	2	EF025736.1	4 e−18	complement component 2/factor B variant 2 (C2/Bf) mRNA, complete cds; *Carcinoscorpius rotundicauda*
s_355027	21853	33	1	XM_002433461.1	0	spectrin alpha chain, putative, mRNA; *Ixodes scapularis*
s_348946	9054	31	3	XM_003702189.1	0	PREDICTED: homeodomain-interacting protein kinase 2-like, transcript variant 2 (LOC100875514), mRNA; *Megachile rotundata*
s_351023	7100	27	1	XM_005093370.1	3 e−08	PREDICTED: nesprin-1-like (LOC101864208), transcript variant X1, mRNA; *Aplysia californica*
s_354499	14549	24	2	XM_002737476.1	0	PREDICTED: Ccar1 protein-like (LOC100372630), mRNA; *Saccoglossus kowalevskii*
s_350235	5615	20	2	AB533506.1	0	EfPiwiB mRNA for piwi, complete cds; *Ephydatia fluviatilis*
s_353164	21201	16	1	XM_005168061.1	3 e−127	PREDICTED: sarcolemma associated protein a (slmapa), transcript variant X2, mRNA; *Danio rerio*
s_353747	14270	16	1	XM_002595709.1	1 e−155	hypothetical protein, mRNA; *Branchiostoma floridae*
s_356524	10745	13	1			No significant hits.
s_355281	23014	13	1	XM_002734134.1	2 e−34	PREDICTED: predicted protein-like (LOC100373712), mRNA; *Saccoglossus kowalevskii*
s_348948	3558	13	2	XR_036743.2	3 e−09	PREDICTED: hypothetical LOC100116128 (LOC100116128), miscRNA; *Nasonia vitripennis*
s_351714	17222	7	2	XM_005315584.1	2 e−53	PREDICTED: capicua transcriptional repressor (CIC), mRNA; *Chrysemys picta* bellii
s_347557	7703	6	3	NM_001204684.1	3 e−154	Mnk (LOC100533401), mRNA; *Aplysia californica*
s_351573	8900	6	1	XM_005291114.1	2 e−38	PREDICTED: uncharacterized LOC101934397 (LOC101934397), mRNA; *Chrysemys picta* bellii
s_353576	16180	5	3	XM_005108846.1	4 e−86	PREDICTED: collagen alpha-6(VI) chain-like (LOC101853123), transcript variant X1, mRNA; *Aplysia californica*
s_354062	14578	3	1	CU207259.10	2 e−04	DNA sequence from clone DKEY-165K4, complete sequence; *Danio rerio*
s_352231	17755	3	1	XM_005108923.1	3 e−110	hamartin-like (LOC101851611), mRNA; *Aplysia californica*

transcriptome; N50 = 859 bp), with 144,014 components (the *Trinity* equivalent to gene regions) in the final assembly. Of the 248 core eukaryotic genes, 93.6% were at least partially present in the assembly, and 83.1% were considered complete by CEGMA criteria. Overall, 20,887 (14.5%) components could be functionally annotated: 17,692 with Pfam domains and 16,526 components with InterPro annotations (13,331 share both). Only 100 (0.07% of the total components) were flagged as microbial, and eight components were identified as mitochondrial, based on a search against a complete polychaete mitochondrial gene set (Won et al. 2013).

We identified 189,574 gene-based SNPs, including 27,295 SNPs that were genotyped confidently in all eight worms. Of this set of SNPs, 240 occur in the eight mitochondrial components. As these SNPs are in known linkage disequilibrium and are not diploid, they were removed from further analysis. We therefore use a set of 27,055 high-quality SNPs for our analysis. These SNPs occur in 2,815 components, which is only 2% of the total components. Most of the components containing high-quality SNPs are annotated (2,411), meaning that 86% of components containing high-quality SNPs are in the 12% of the components that have annotations. Of the 15,056

non-singletons in this set, 9,164 are shared, with both alleles observed in both populations. A further 3,272 SNPs are private to Bayonne while 2,589 are private to Long Beach. Only 31 SNPs are fixed differences, with all four Bayonne worms homozygous for one allele and all four Long Beach worms homozygous for the other. These fixed differences occur in only 19 transcriptome components. Most of these fixed SNP-containing components were annotated (Table 1). If we consider sites genotyped in at least three worms per population (i.e., including sites with some missing data), 56,587 of 103,571 SNPs are non-singletons, and of these, we observe 34,581 shared variants, 21,851 private SNPs, and 155 fixed differences.

An exact permutation test for the 15,056-SNP dataset indicates that the data are inconsistent with panmixia (Fig. 2a). The permuted data sets show a greater distribution of likelihoods than the simulations, which is likely due to effects of linkage disequilibrium in the dataset. The likelihood calculated from our actual data set is far smaller than either the simulated or permuted datasets. This demonstrates that the two populations are diverged more than we would expect under random mating and confirms the reliability of our transcript-based SNP genotypes. Although the majority of SNPs in our dataset are shared polymorphisms (9,164), the expected number of shared variants under panmixia is even greater (mean 12,060), and the expected number of fixed differences is less than one (Fig. 2b).

The best-fitting demographic model for our data is an IM model with $\theta = 3{,}365$ and the parameters shown in Table 2. Note that our two samples have been isolated on separate coasts for only a century, and that for our purposes Bayonne and Long Beach stand as representatives of the deeper history relating planktotrophs and lecithotrophs.

Figure 3 shows the fit of this model to our full dataset and a randomly sub-sampled dataset with one SNP per transcriptome component. These parameter estimates provide a rough outline of the demographic history of our populations, which additional individuals per population may clarify. Of specific interest is the estimate that Bayonne has an eightfold larger effective population size than Long Beach, perhaps due to the bottleneck associated with the very recent introduction of *S. benedicti* to the California coast, and that long-term historical gene flow between these populations has been asymmetric with 2.7× more gene flow occurring from the lecithotrophic to planktotrophic population. It is important to emphasize that this migration estimate reflects long-term historical gene flow, as these two populations are only recently isolated on different coasts. Our model suggests that separation

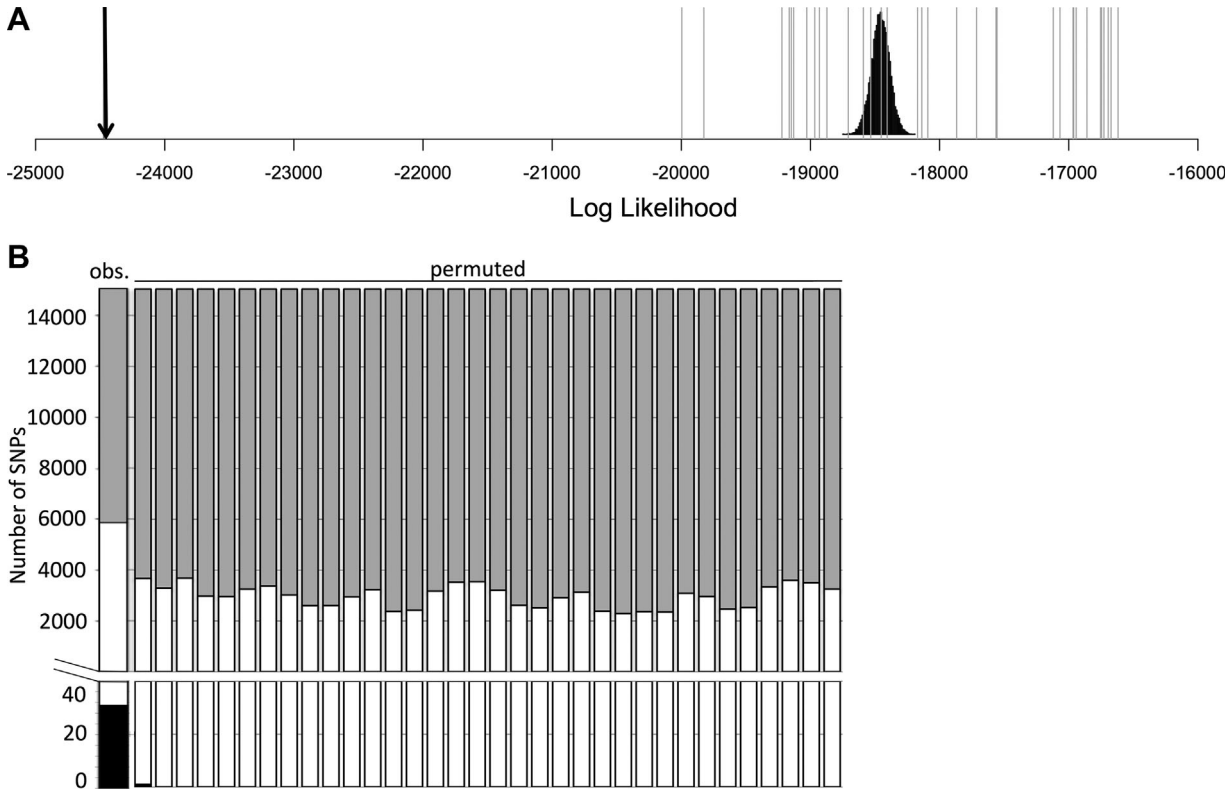

Fig. 2. A: The genotype configuration likelihood under panmixia for our observed data (arrow) and all permutations of the eight individuals (grey bars). The black histogram is the log likelihood of the data simulated under the null hypothesis of panmixia with independence among loci. B: SNP distribution across the two populations. The proportion of shared (gray), private (white) and fixed (black) SNPs are shown for the observed data and each of the 34 permutations of the data.

Table 2. Best fitting IM model parameters from ∂a∂i

Parameter	Estimate	CI	Unit
Fraction of ancestral population that splits into Bayonne	0.619	0.22–1.0	$N_{eBay}/N_{eAncestral}$
Relative size of Bayonne to ancestral population	8.06	4.64–14.50	$N_{eBay}/N_{eAncestral}$
Relative size of Long Beach to ancestral population	1.04	0.45–1.61	$N_{eLB}/N_{eAncestral}$
Time since divergence	4.68	1.46–7.38	$2N_e$ generations
Migration Bayonne–Long Beach	0.65	0.13–1.21	$2N^em$
Migration Long Beach–Bayonne	1.76	0.64–2.66	$2N^em$

of these populations occurred in the distant past, ∼2.3*N generations ago, but gene flow between populations has been occurring with ∼0.32 (Bayonne→Long Beach) and ∼0.88 (Long Beach→Bayonne) effective migrants per generation, substantially homogenizing the populations.

DISCUSSION

We use comparative transcriptome sequencing to investigate the gene-based SNP differences between worms of contrasting larval types from two populations. Our sequencing coverage surpasses any previous genomic information for *S. benedicti,* and we have sufficient coverage of each individual to recover a large number of conservatively-called SNPs.

We find that there are more fixed differences and fewer shared polymorphisms that we would expect under a panmictic population, demonstrating that there is genetic structure between the two populations. This, of course, is not surprising given that these individuals are from two distinct geographic populations. Absent more extensive sampling, the geographic distribution of larval types, with populations that are predominately if not exclusively one type (described in Zakas and Wares 2012) confounds our ability to disentangle geographic and life-history differences.

Across thousands of highly expressed transcripts and SNPs there is modest differentiation, at best, despite large geographic differences. Among high confidence non-singleton SNPs, most are shared polymorphisms and very few, about 0.2%, are fixed differences. Considering we use a small sample of only eight worms, adding individuals will likely reduce this number even further. This striking lack of differentiation across the transcriptome demonstrates that gene flow and/or recent shared ancestry between larval types is extensive, and importantly, the severe larval dimorphism in *S. benedicti* is not occurring due to fixed allelic differences in large numbers of transcripts. Although our survey is limited to transcribed regions, we would likely detect differentiation due to selection at tightly linked regulatory regions, particularly in the presence of or large blocks of linkage-disequilibrium maintained by recombition-suppressing inversions.

While our results reject a panmictic null model, as we would expect, our demographic modeling suggests that there has been extensive gene flow between populations. When interpreting the model results, it is important to note that while these populations are extremely unlikely to be currently sharing migrants, the Long Beach population is a recent introduction from the East Coast. Historically there could be migration between the founding population of Long Beach and Bayonne. In fact, the model suggests that this historical effective migration was fairly substantial. It is also interesting that the predicted migration is asymmetrical, with greater gene flow from the lecithotrophic to the planktotrophic population. While this warrants further investigation in mating studies and more extensive population sampling, this result may suggest biased migration between populations.

The fit of the diffusion approximation model to our data illustrates important insights into the genetic architecture of life-history traits. The model is fitting the demographic differences between the populations that we expect under historical divergence with migration. Overall the model is a good fit for the shared polymorphisms, with no overall trend of over or underrepresentation across categories. The residuals between our model and the data illustrate the regions of the SFS that are not a fit to demographic predictions, and are likely due to variants in linkage disequilibrium. When the data is corrected using subsamples with no known linkage disequilibrium, the model effectively explains the data with very little residual differences. (Fig. 3b). We conclude that the number of fixed differences and private alleles in our data set are not different from what we would expect under this IM model, illustrating that there is not a large amount of differentiation between these two populations despite contrasting development modes.

These results begin to inform our understanding of contrasting larval morphologies, and life histories, in a single species. Given the set of heritable differences in larval phenotypes for these two populations (Levin et al. 1991; Zakas and Rockman 2014), we might suspect that a considerable amount of the transcriptome would harbor divergent coding regions and fixed SNP differences. However, our best-fitting model, in addition to previous work (Zakas and Wares 2012), demonstrates there is gene flow between populations differing in larval type. While the exact

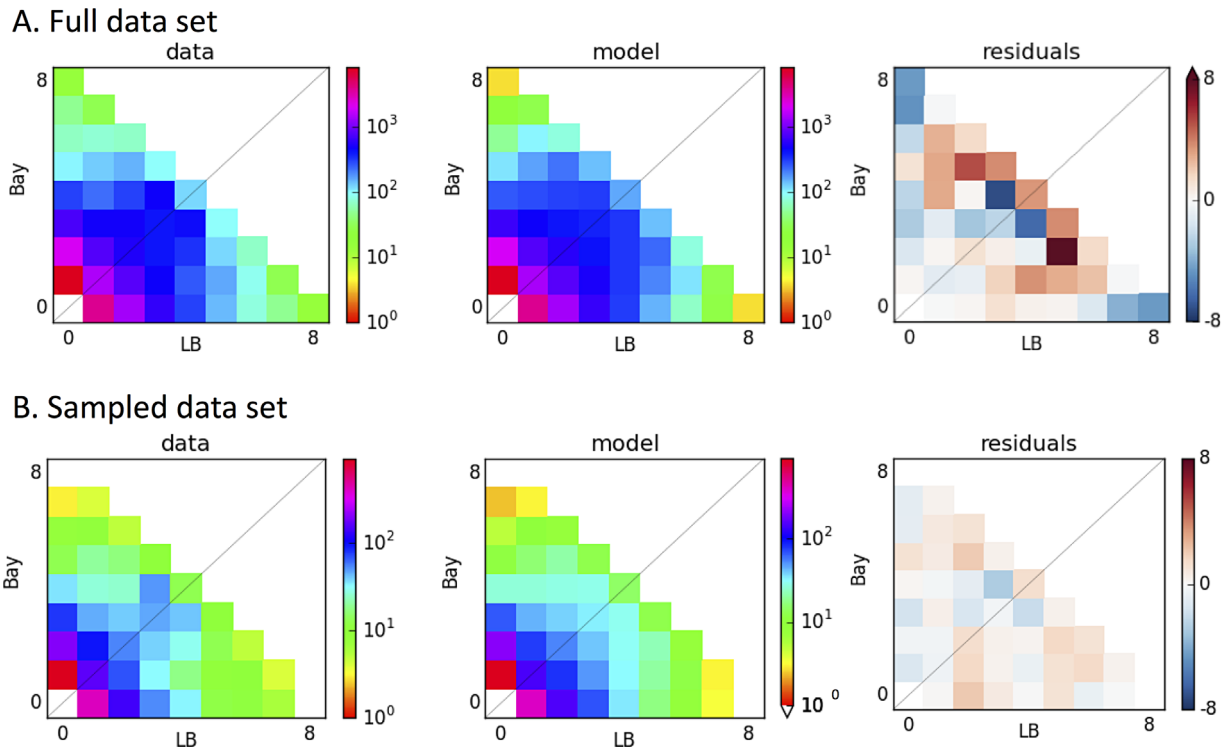

Fig. 3. Fit of the best model to the SFS using ∂a∂i. A: The folded SFS heat map for 8 haplotypes in 2 populations for the full dataset and the best-fitting model, as well as the heat map of the normalized residuals between the model and data. B: The folded SFS heat map for a randomly sampled dataset with no known linkage disequilibrium. The heat-map of the residuals shows the dataset without linkage disequilibrium is a better fit to the model than the full dataset.

extent and direction of this gene flow is hard to determine from just a few individuals per population, there is nonetheless extensive historical gene flow. Therefore models of divergence with gene flow are applicable to the maintenance of this strategy. The genetic architecture of divergence with ongoing gene flow usually involves discrete genomic regions (which may be isolated or distributed throughout the genome) that contribute to differentiation (Peichel et al. 2001; Emelianov et al. 2004; Kulathinal et al. 2009; Nosil et al. 2012; Martin et al. 2013; Marchinko et al. 2014). For example, in cases of ecological speciation, regions of restricted recombination such as centromeres, inversions (Noor et al. 2001), or mosaic regions of genomic isolation will maintain fixed haplotypes that contribute to phenotypic differentiation (for example, Via et al. 2012, reviewed in Pinho and Hey 2010). We have not recovered evidence of divergent genomic regions so far, but it is unlikley that we would be able to detect genomic islands of divergence in transcriptome data (Cruickshank and Hahn 2014), warranting more extensive sampling of the both individuals and non-coding genomic markers.

Alternatively, this pattern of geographic differentiation across populations with limited evidence of divergence between larval modes could indicate recurrent ecological diversification at the population-level. For example, *Littorina saxatilis* (Panova et al. 2006; Butlin et al. 2014) has morphologically divergent ecotypes, resulting in a similar pattern to that of *S. benedicti*. In *L. saxatilis*, ecotypes appear to have repeatedly evolved in sympatry, and the genetic differences contributing to ecotype are specific to subpopulations. While it is possible a similar ecologically driven mechanism could be responsible for the geographic pattern of offspring type in *S. benedicti*, a major distinction is the absence of known assortative mating or reproductive barriers in *S. benedicti* that would be maintaining the divergent types (Schulze et al. 2000; Zakas and Rockman 2014).

While we analyze only a portion of the genome here, we find no evidence that these two larval morphs are incipient species, leaving the mechanism responsible for maintaining this genetically based dimorphism in the presence of gene flow undetermined. Understanding the genetic mechanisms that shape development mode remains critical, and we anticipate that future work on the emerging *S. benedicti* model will reveal the genetic and regulatory mechanisms driving poecilogony and developmental divergence.

ACKNOWLEDGMENTS

We would like to thank B. Pernet for providing animals from Long Beach, J. Nicodemus for preparing sequencing libraries, A. Paaby, V. Cattani, M. Bernstein, L. Noble and T. Kaur for feedback on the manuscript; L. Noble and J.D. Robinson for bioinformatic expertise, and J. Deutscher for larval imaging. This work is supported by funds

from New York University, the Zegar Family Foundation, an NIH NRSA Postdoctoral Fellowship # 1F32GM108396, and NSF grant IOS-1350926.
Transcriptome sequences are available in the NCBI database, BioProject ID PRJNA283543.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403–410.

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., and Kabir, S, et al. 2014. Genetics of ecological divergence during speciation. *Nature* 511: 307–311.

Bridges, T. 1993. Reproductive investment in four developmental morphs of Streblospio (Polychaeta: spionidae). *Biol Bull* 184: 144–152.

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., et al. 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* 68: 935–949.

J. T. Carlton , 1979. History, biogeography and ecology of the introduced marine and estuarine invertebrates of the Pacific Coast of North America. PhD. Thesis., University of California, Davis.

Charlesworth, B., Nordborg, M., and Charlesworth, D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* 70: 155–174.

Cruickshank, T. E., and Hahn, M. W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23: 3133–3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., and DePristo, M. A. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Emelianov, I., Frantisek, M., and Mallet, J. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proc Roy Soc B* 271: 97–105.

Feder, J. L., and Nosil, P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64: 1729–1747.

Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., and Hotz, HR, et al. 2008. The Pfam protein families database. *Nucleic Acids Res* 36: D281–D288.

Gibson, G., MacDonald, K., and Dufton, M. 2010. Morphogenesis and phenotypic divergence in two developmental morphs of *Streblospio benedicti* (Annelid Spionidae). *Invert Biol* 129: 328–343.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., and Amit, I. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* 29: 644–652.

Gutenkunst, R. N., Herandez, R. D., Williamson, S. H., and Bustamante, C. D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5: e1000695.

Kulathinal, R. J., Stevison, L. S., and Noor, M. A. F. 2009. The genomics of speciation in *Drosophila:* Diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet* 5: e1000550.

Levin, L. A. and Bridges, T. S. 1994. Control and consequences of alternative developmental modes in a poecilogonous polychaete. *Amer Zool* 34: 323–332.

Levin, L. A. and Creed, E. L. 1986. Effect of temperature and food availability on reproductive responses of *Streblospio benedicti* (Polychaeta: Spionidae) with planktotrophic or lecithotrophic development. *Mar Biol* 92: 103–113.

Levin, L. A. and Huggett, D. V. 1990. Implications of alternative developmental reproductive modes for seasonality and demography in an estuarine polychaete. *Ecology* 71: 2191–2208.

Levin, L. A., Zhu, J., and Creed, E. 1991. The genetic basis of life-history characters in a polychaete exhibiting planktotrophy and lecithotrophy. *Evolution* 45: 380–397.

Levin, L. A. 1984. Multiple patterns of development in *Streblospio benedicti* Webster (Spionidae) from three coasts of North America. *Biol Bull* 166: 494–508.

Marchinko, K. B., Matthews, B., Arnegard, M. E., Rogers, S. M., and Schluter, D. 2014. Maintenance of a genetic polymorphism with disruptive natural selection in stickleback. *Curr Biol* 24: 1289–1292.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., and Simpson, F. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res* 23: 1817–1828.

McCain, E. R. 2008. Poecilogony as a tool for understanding speciation: Early development of *Streblospio benedicti* and *Streblospio gynobranchiata* (Polychaeta: Spionidae). *Invert Repro Dev* 51: 91–101.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., and Kernytsky, A. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303.

Noor, M. A., Grams, K. L., Bertucci, L. A., and Reiland., J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* 98: 12084–12088.

Nosil, P., Gompert, Z., Farkas, T. E., Comeault, A. A., Feder, J. L., Buerkle, C. A., and Parchman, T. L. 2012. Genomic consequences of multiple speciation processes in a stick insect. *Proc Roy Soc B* 279: 5058–5065.

Panova, M., Hollander, J., and Johannesson, K. 2006. Site-specific genetic divergence in parallel hybrid zoned suggests nonallopatric evolution of reproductive barriers. *Mol Ecol* 13: 4021–4031.

Parra, G., Bradnam, K., and Korf., I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061–1067.

Peichel, C. L., Nereng, K. S., Ohgi, K. A., Cole, B. L. E., Colosimo, P. F., and Buerkle, C. A., et al. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* 414: 901–905.

Pernet, B. and McHugh, D. 2010. Evolutionary changes in the timing of gut morphogenesis in larvae of the marine annelid *Streblospio benedicti*. *Evol Dev* 12: 618–627.

Pinho, C. and Hey, J. 2010. Divergence with gene flow: Models and data. *Annu Rev Ecol Evol Syst* 41: 215–230.

R Core Development Team. 2011. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raff, R. A. (1996) *The shape of life: genes, development, and the evolution of animal form*. Chicago: The University of Chicago Press.

Rockman, M. 2012. Patterns of nuclear genetic variation in the poecilogonous polychaete *Streblospio benedicti*. *Integr Comp Biol* 52: 173–180.

Schulze, S. R., Rice, S. A., Simon, J. L., and Karl, S. A. 2000. Evolution of poecilogony and the biogeography of North American populations of the polychaete *Streblospio*. *Evolution* 54: 1247–1259.

Via, S., Conte, G., Mason-Foley, C., and Mills, K. 2012. Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Mol Ecol* 21: 5546–5560.

Won, E. J., Rhee, J. S., Shin, K. H., and Lee, J. S. 2013. Complete mitochondrial genome of the marine polychaete *Perinereis nuntia* (Polychaeta, Nereididae). *Mitochondr DNA* 24: 342–343.

Wood, D. E. and Salzberg, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15: R46.

Yeaman, S. and Whitlock, M. C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* 65: 1897–1911.

Zakas, C. and Wares, J. P. 2012. Consequences of a poecilogonous life-history for genetic structure in coastal populations of the polychaete *Streblospio benedicti*. *Mol Ecol* 21: 5447–5460.

Zakas, C., Schult, N., McHugh, D., Jones, K., and Wares, J. P. 2012. Transcriptome analysis and SNP development can resolve population differentiation of *Streblospio benedicti*, a developmentally dimorphic marine annelid. *PLoS ONE* 7: e31613.

Zakas, C. and Rockman, M. V. 2014 Dimorphic development in *Streblospio benedicti*: genetic analysis of morphological differences between larval types. *Int J Dev Biol*. doi: 10.1387/ijdb.140088mr.

Zdobnov, E. M. and Apweiler, R. 2001. InterProScan- an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.