

# Insight Into Genomic Changes Accompanying Divergence: Genetic Linkage Maps and Synteny of *Lucania goodei* and *L. parva* Reveal a Robertsonian Fusion

Emma L. Berdan,<sup>\*1,2,3</sup> Genevieve M. Kozak,<sup>\*1</sup> Ray Ming,<sup>†</sup> A. Lane Rayburn,<sup>‡</sup> Ryan Kiehart,<sup>§</sup> and Rebecca C. Fuller<sup>\*</sup>

<sup>\*</sup>Department of Animal Biology, University of Illinois, Champaign, Illinois 61820, <sup>†</sup>Department of Plant Biology, and <sup>‡</sup>Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801, and <sup>§</sup>Department of Biology, Ursinus College, Collegeville, Pennsylvania 19426

**ABSTRACT** Linkage maps are important tools in evolutionary genetics and in studies of speciation. We performed a karyotyping study and constructed high-density linkage maps for two closely related killifish species, *Lucania parva* and *L. goodei*, that differ in salinity tolerance and still hybridize in their contact zone in Florida. Using SNPs from orthologous EST contigs, we compared synteny between the two species to determine how genomic architecture has shifted with divergence. Karyotyping revealed that *L. goodei* possesses 24 acrocentric chromosomes (1N) whereas *L. parva* possesses 23 chromosomes (1N), one of which is a large metacentric chromosome. Likewise, high-density single-nucleotide polymorphism–based linkage maps indicated 24 linkage groups for *L. goodei* and 23 linkage groups for *L. parva*. Synteny mapping revealed two linkage groups in *L. goodei* that were highly syntenic with the largest linkage group in *L. parva*. Together, this evidence points to the largest linkage group in *L. parva* being the result of a chromosomal fusion. We further compared synteny between *Lucania* with the genome of a more distant teleost relative medaka (*Oryzias latipes*) and found good conservation of synteny at the chromosomal level. Each *Lucania* LG had a single best match with each medaka chromosome. These results provide the groundwork for future studies on the genetic architecture of reproductive isolation and salinity tolerance in *Lucania* and other Fundulidae.

## KEYWORDS

synteny  
Robertsonian  
fusion  
chromosomal  
rearrangement  
linkage map  
speciation  
EST-based SNPs  
fundulidae

Copyright © 2014 Berdan *et al.*

doi: 10.1534/g3.114.012096

Manuscript received February 25, 2014; accepted for publication May 20, 2014; published Early Online June 3, 2014.

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.114.012096/-/DC1>

The population-specific Illumina transcriptome sequences are archived in GenBank (bio-project ID: PRJNA215087). Source and probe sequences from the Illumina Infinium chip are accessible on Dryad (<http://datadryad.org/resource/doi:10.5061/dryad.hv75h/9>).

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Present address: Museum für Naturkunde, Leibniz-Institut für Evolutions-und Biodiversitätsforschung, Berlin, Germany.

<sup>3</sup>Corresponding author: Museum für Naturkunde, Leibniz-Institut für Evolutions-und Biodiversitätsforschung, Invalidenstraße 43, 10115 Berlin, Germany.

E-mail: emma.berdan@mfn-berlin.de

Species-specific linkage maps are critical to understanding genomic architecture and how it differs between species. Linkage maps allow the exploration of genotype-phenotype relationships through quantitative trait loci mapping (Falconer and Mackay 1996; for examples, see Peichel *et al.* 2001; Tripathi *et al.* 2009) and the comparison of genomic architecture between species via synteny mapping (Prince *et al.* 1993; Backstrom *et al.* 2008; Muchero *et al.* 2009; Lucas *et al.* 2011; McGraw *et al.* 2011). However, the majority of previous studies of genome-wide synteny have used pairs of species that are distantly related (*i.e.*, from different orders or families: Stapley *et al.* 2008; Jaari *et al.* 2009; Foley *et al.* 2011; McGraw *et al.* 2011). It is less common for comparisons of synteny to be made within the same genus (but see Ming *et al.* 1998; Rogers *et al.* 2007; Lubieniecki *et al.* 2010; Lee *et al.* 2011; Timusk *et al.* 2011; Naish *et al.* 2013). High-density linkage maps are now possible with the advent of high-throughput sequencing and have the potential to facilitate fine-scale comparisons of synteny between closely related species. These comparisons are needed to

determine how genome structure diverges and potentially contributes to speciation.

The problem for maintaining species boundaries in areas of sympatry is that gene flow between species and recombination in hybrids should homogenize species-specific traits and break down reproductive isolating barriers. Genomic rearrangements, such as chromosomal fusions, inversions, or deletions, can potentially facilitate the maintenance of reproductive isolating barriers because they reduce recombination in portions of the genome (Rieseberg 2001; Butlin 2005; Faria and Navarro 2010). Data from a wide range of taxa support the theory that genes conferring reproductive isolation between sympatric species often are localized to rearranged areas of the genome (Rieseberg *et al.* 1999; Noor *et al.* 2001; Coluzzi *et al.* 2002; Feder *et al.* 2003; Kitano *et al.* 2009).

One way to study the role of genomic rearrangements in speciation is to compare synteny between species that hybridize at low levels. We do this in *Lucania*, a genus that is becoming a model system for ecological speciation research. The bluefin killifish (*Lucania goodei*) and the rainwater killifish (*L. parva*) are two closely related species that differ radically in their salinity tolerance (Duggins *et al.* 1983; Whitehead 2010). *Lucania goodei* is found primarily in freshwater habitats, whereas *L. parva* is euryhaline and can be found in fresh, brackish, and marine habitats (Lee *et al.* 1980). Survival at various life stages differs between *Lucania* species in different salinities (Fuller *et al.* 2007; Fuller 2008). *Lucania goodei* and *L. parva* also show divergence in sequence and expression of a number of salinity tolerance genes (Berdan and Fuller 2012; Kozak *et al.* 2014). Despite these ecological differences between species, there is evidence for low levels of ongoing hybridization in sympatric populations. Sympatric populations exist in the Atlantic and Gulf coastal waterways of Florida (Fuller and Noa 2008). Hubbs *et al.* (1943) found hybrids based on morphological characters in one population, and analysis of mtDNA suggests recent gene flow between the two species in multiple river drainages in Florida (R. C. Fuller, unpublished data). We hypothesized that differences in genome structure might contribute to speciation in *Lucania*. This hypothesis was sparked by an unsupported comment made in an older study; Uyeno and Miller (1971) stated that chromosome number differs between *L. goodei* and *L. parva* but did not provide any evidence to support this statement.

The purpose of this study was to determine whether *L. goodei* and *L. parva* differ in genomic architecture. To do this, we (1) karyotyped both species and (2) produced two high-density single-nucleotide polymorphism (SNP)-based linkage maps. Transcriptome sequencing and high-throughput SNP genotyping were performed in both species to create the linkage maps. These data allowed us to determine (1) the number of chromosomes possessed by each species, (2) whether a fusion/fission event had occurred, and (3) patterns of synteny between the two species. We analyzed synteny at the linkage group (LG) level and at the level of marker order to determine whether large or small-scale genomic rearrangements have occurred during divergence (Gale and Devos 1998; Ming *et al.* 1998; Wang *et al.* 2010). We further compared synteny in *Lucania* with the most closely related species with a sequenced genome, medaka (*Oryzias latipes*) to ask how *Lucania* genomic architecture corresponds to that of other teleost fish (Kasahara *et al.* 2007). These linkage maps will enable future studies to ask whether the areas of the genome contributing to salinity tolerance also are implicated in reproductive isolation and whether these traits map to genomic rearrangements in *Lucania*.

## MATERIALS AND METHODS

### Karyotyping

Somatic karyotypes were determined for both *L. goodei* and *L. parva* using metaphase spreads. For each species, animals of both sexes from

multiple populations were used. Details on the karyotyping methods can be found in Supporting Information, File S1. In summary, individuals were injected intraperitoneally at 0 hr with a phytohemagglutinin solution to stimulate mitosis, injected at 24 hr with 1% colchicine solution, and then killed at 26 hr using MS-222. The gills were removed, placed in chilled distilled water to allow the cells to swell (30 min), then fixed in a 3:1 methanol/glacial acetic acid mixture (30 min). The gills were dabbed on the slides to break the cells and release the chromosomes. Slides were cleared, dried, stained with 5% Giemsa solution (Electron Microscopy Sciences, Hatfield, PA), mounted with Permount, and visualized using a compound microscope.

### Comparative genetic linkage maps: overview

Genetic linkage maps were created separately for both *L. goodei* and *L. parva* from F<sub>2</sub> mapping crosses between two geographically isolated populations. The key to comparing the genetic linkage maps between the two species was to use SNP markers from orthologous expressed sequence tags (ESTs; from 454 and Illumina sequencing) that were expressed in both species (Figure 1). F<sub>2</sub> offspring and F<sub>1</sub> parents were genotyped using an Illumina Infinium Bead Chip custom designed for *Lucania*.

### Mapping crosses

For each species, crosses were set up between geographically and ecologically divergent populations. For *Lucania goodei*, these populations were Upper Bridge from the spring-fed Wakulla River (Wakulla County, Florida) and a swamp population in the Everglades (Broward County, Florida). For *Lucania parva*, the populations were Indian River Lagoon, an Atlantic coast population where salinity is typically 35 ppt (Brevard County, Florida) and Pecos River, a freshwater inland river in Texas (Pecos-Crockett County border, Texas). Between population breeding pairs were established in both hybrid cross directions. Cross designs for the creation of the F<sub>1</sub> parents are described for *L. goodei* in Fuller *et al.* (2010) and for *L. parva* in Kozak *et al.* (2012). F<sub>1</sub> hybrid offspring were then raised to adulthood and paired with unrelated F<sub>1</sub> hybrid individuals to create F<sub>2</sub> offspring. For *L. parva*, we created and genotyped 161 F<sub>2</sub> offspring from 8 F<sub>2</sub> families and 16 F<sub>1</sub> parents (parents from 11 F<sub>1</sub> families: 4 Indian River female × Pecos male crosses; 7 Pecos female × Indian River male crosses; see Kozak *et al.* 2012). For *L. goodei*, we created and genotyped 303 F<sub>2</sub> offspring

- Step 1. Isolate RNA from gills, fins, eyes, brains, ovaries, and testes for males and females from 4 populations (2 *L. goodei* populations and 2 *L. parva* populations) using 10 individuals from each population.
- Step 2. Create pooled samples for each population using the same amount of RNA from each individual.
- Step 3. Prepare cDNA libraries uniquely barcoded by population.
- Step 4. Sequence 100-bp paired-end libraries with Illumina.
- Step 5. Trim ends and check sequence quality.
- Step 6. Use contigs from previous 454 sequencing of *L. goodei* (Fuller and Claricoates 2011) as a reference for assembling all Illumina sequences.
- Step 7. Assemble each of 4 populations separately (*L. parva*: Pecos River and Indian River; *L. goodei*: Upper Bridge and Everglades) using Novoalign.
- Step 8. Export the assemblies to MAQ software. Find diagnostic, population-specific SNPs using MAQ's SNP caller.
- Step 9. Assess SNP quality and choose SNPs from contigs that have all types of SNPs (*L. goodei* population-specific, *L. parva* population-specific, between species).
- Step 10. Genotype mapping families at all SNPs using custom designed Illumina Infinium Bead Chip.
- Step 11. Create linkage maps for each species in JoinMap.

**Figure 1** Procedures used to generate single-nucleotide polymorphisms (SNPs) from orthologous expressed sequence tags (ESTs) for linkage maps.

from 14 F<sub>2</sub> families and 28 F<sub>1</sub> parents (21 F<sub>1</sub> families: 11 Upper Bridge female × Everglades male crosses; 10 Everglades female × Upper Bridge male crosses).

F<sub>2</sub> eggs were collected and raised in freshwater with dilute methylene blue (an antifungicide). Fry were fed newly hatched *Artemia* and raised to 1 month postfertilization. The fry were then killed in MS-222, preserved in ethanol, and stored at -80°. F<sub>1</sub> parents and F<sub>0</sub> grandparents were preserved in ethanol. F<sub>2</sub> family sizes ranged from 15 to 24 fry.

### Creation of population-specific EST libraries

RNA was extracted from five males and five females from the two *L. goodei* populations (Upper Bridge and Everglades) and the two *L. parva* populations (Indian River and Pecos River). Fish were killed in MS-222. Tissue samples were taken from the gills (1–2 arches), dorsal fins, eyes, brain, and the gonads (ovaries or testes). RNA was extracted using a protocol modified from Carleton (2011; see File S1). A pooled sample containing RNA from all tissues was created for each population that contained equal amounts of RNA from all individuals. The RNA was submitted to the Keck Center for Comparative and Functional Genomics at the University of Illinois for creation of cDNA libraries and sequenced using Illumina HiSeq 2000 (see File S1). The two populations from *L. goodei* were uniquely barcoded and run on a single lane of Illumina HiSeq for 100-bp paired-end sequencing. Indian River *L. parva* and Pecos River *L. parva* were run on separate lanes. Average quality scores were above 20 for all cycles. Samples produced from 5 to 10 billion bases of data (reads per end: Upper Bridge = 29,661,140, Everglades = 26,235,855, Pecos River = 55,535,778, Indian River = 53,061,233). The population specific Illumina transcriptome sequences are archived in Genbank (bio-project ID: PRJNA215087).

In addition, 454 sequencing was done on pooled samples of RNA from five *L. goodei* populations across Florida (1: Upper Bridge Wakulla River, Wakulla, Co., FL; 2: St. Mark's National Wildlife Refuge Gambo Bayou, Wakulla, Co., FL; 3: 26-Mile Bend, Everglades, Broward Co., FL; 4: Rum Island Park, Santa Fe River, Columbia Co., FL; and 5: Delk's Bluff Bridge, Oklawaha River, Marion Co., FL). These sequences were used to construct a reference upon which subsequent assembly was based (see section: *Assembly and alignment*; reference available in Dryad accompanying this paper). Tissue samples were taken from the gills, fins, eyes, brain, and gonads (ovaries or testes). Additional details on the 454 project can be found in Fuller and Claricoates (2011).

### Assembly and alignment

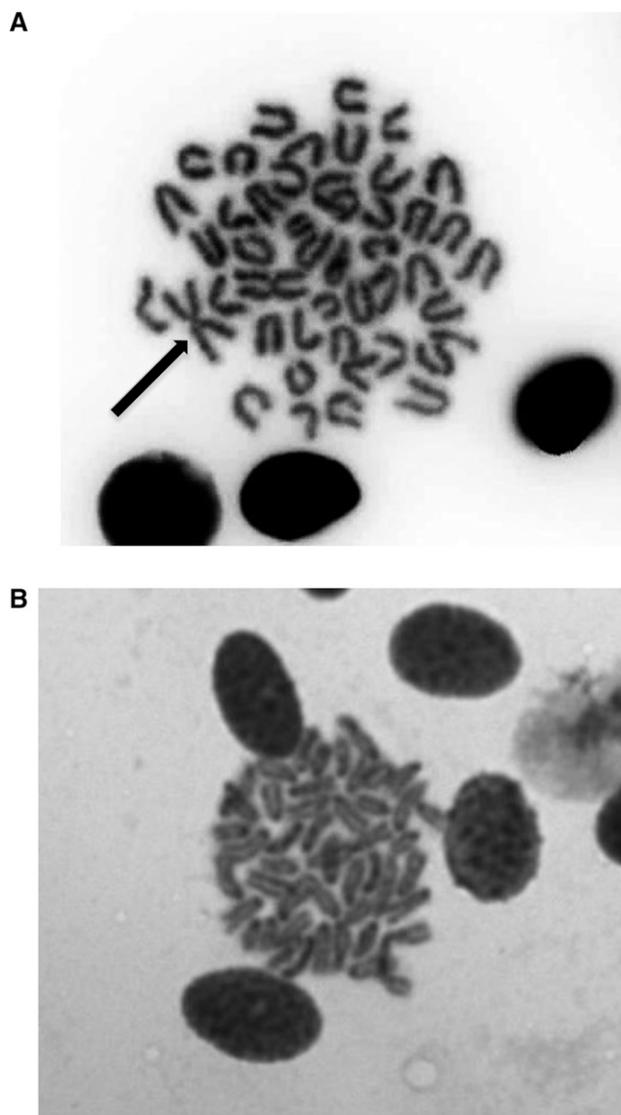
*Lucania goodei* 454 sequences were used as a reference for the assembly. For the reference, contigs were assembled using Newbler assembler (454 Life Sciences, Branford, CT). Assembly parameters were as follows: the minimum contig length was set at 200 bp, the minimum overlap length was 60 bp, and the minimum overlap identity was 95%. A total of 29,838 contigs were generated by the Newbler assembler. Using the *L. goodei* 454 contigs as a reference anchored the analysis of shorter Illumina sequences and allowed identification of contigs that contained multiple SNPs. The goal was to find contigs containing a SNP that was diagnostic for the two *L. goodei* populations as well as a SNP that was diagnostic for the two *L. parva* populations. This method may have missed *L. parva* contigs that were not expressed in *L. goodei*, but these contigs are uninformative for the comparison of linkage maps. Illumina sequences were trimmed to 75 bp in length and aligned against the reference using Novoalign (Novocraft Technologies; [www.novocraft.com](http://www.novocraft.com)). Alignment parameters were as follows: the maximum alignment score acceptable for a best alignment was set

at 45; the gap extend penalty was set at 10; and the number of good quality bases for an acceptable read was set at 50.

### SNP selection

The alignments were exported to MAQ (*i.e.*, Mapping and Assembling with Quality) software (see Fuller and Claricoates 2011 for details) for SNP detection. Diagnostic SNPs for each population were identified using its population pair as a reference. SNPs were considered to be diagnostic when they were identified unambiguously for both populations. There were many more diagnostic SNPs between the two *L. parva* populations than there were between two *L. goodei* populations. To increase the number of SNPs for *L. goodei*, SNPs were used that were fixed in one population but were segregating in the alternate population.

Candidate SNPs were submitted to Illumina for initial evaluation for suitability for the Infinium Genotyping Assay. There were three classes of SNPs: *L. parva*-specific SNPs were SNPs that were segregating within *L. parva*, *L. goodei*-specific SNPs were SNPs that were segregating within *L. goodei*, and between species SNPs that were fixed (or nearly fixed) between the two species. The between species SNPs



**Figure 2** Somatic metaphase/anaphase spread of (A) *Lucania parva* and (B) *L. goodei*. The arrow in (A) indicates the fused chromosome.

■ Table 1 Summary of integrated linkage maps for *L. parva* and *L. goodei*

	<i>L. parva</i>	<i>L. goodei</i>
No. chromosomes	23	24
No. linkage groups	23	24
Map size, cM	605	392
Average linkage group size, cM	26.3	16.3
Markers per cM	1.26	2.41
Total no. markers	766	915
No. individuals	161	303

were designed for another study. The custom Infinium bead chip held probes for 4545 SNPs. Of these, 1497 were candidate SNPs for *L. goodei*, 1369 were candidate SNPs for *L. parva*, and 1679 were candidate between species SNPs. All SNPs were labeled with the ID number of the contig in which they occur. Protein annotations for the linkage map contigs were obtained using blastX searches against telost reference proteomes (Atlantic killifish: *Fundulus heteroclitus*, Japanese medaka: *Oryzias latipes*, three-spined stickleback: *Gasterosteus aculeatus*, guppy: *Poecilia reticulata*) and human complete proteome (Uniprot release 2012\_8 available at: [ftp://ftp.uniprot.org/pub/databases/uniprot/previous\\_releases/release-2012\\_08/uniref/](ftp://ftp.uniprot.org/pub/databases/uniprot/previous_releases/release-2012_08/uniref/)). Only contigs that matched a single protein in each species with a blast score >100 were considered high confidence annotations (Table S1). Proteins that

matched multiple contigs located on different LGs were removed from the annotations.

### DNA extraction and SNP genotyping

DNA was extracted using a modified version of the PureGene (Gentra Systems; [www.gentra.com](http://www.gentra.com)) extraction protocol over 4 days (see File S1). Sample concentration and quality were verified using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). DNA samples were diluted to 75 µg/µL in nuclease-free water and then genotyped at all SNPs using the *Lucania* Illumina Infinium Bead Chip. Source and probe sequences from the chip are accessible on Dryad (<http://datadryad.org/resource/doi:10.5061/dryad.hv75h/9>). Bead chips were scanned using the iScan System (Illumina) at the Keck Center for Comparative and Functional Genomics at The University of Illinois. Illumina GenomeStudio (v2011.1) was used for genotype calls. Cluster positioning was performed separately for *L. goodei* and *L. parva* SNPs (no-call threshold was set to 0.15). Population-specific SNP alleles were verified from genotypes of 16 Upper Bridge *L. goodei* (7 males, 9 females), 17 Everglades *L. goodei* (7 males, 10 females), 6 Indian River *L. parva* (3 males, 3 females), and 5 Pecos River *L. parva* (3 males, 2 females). Genotype data were analyzed separately for each family. For each family, SNPs were removed if either (1) genotypes were homozygous in both parents thereby guaranteeing all offspring were homozygotes, or (2) the genotype was a no-call (*i.e.*, the sample did not run) for one or both parents. Not all SNPs were fixed between

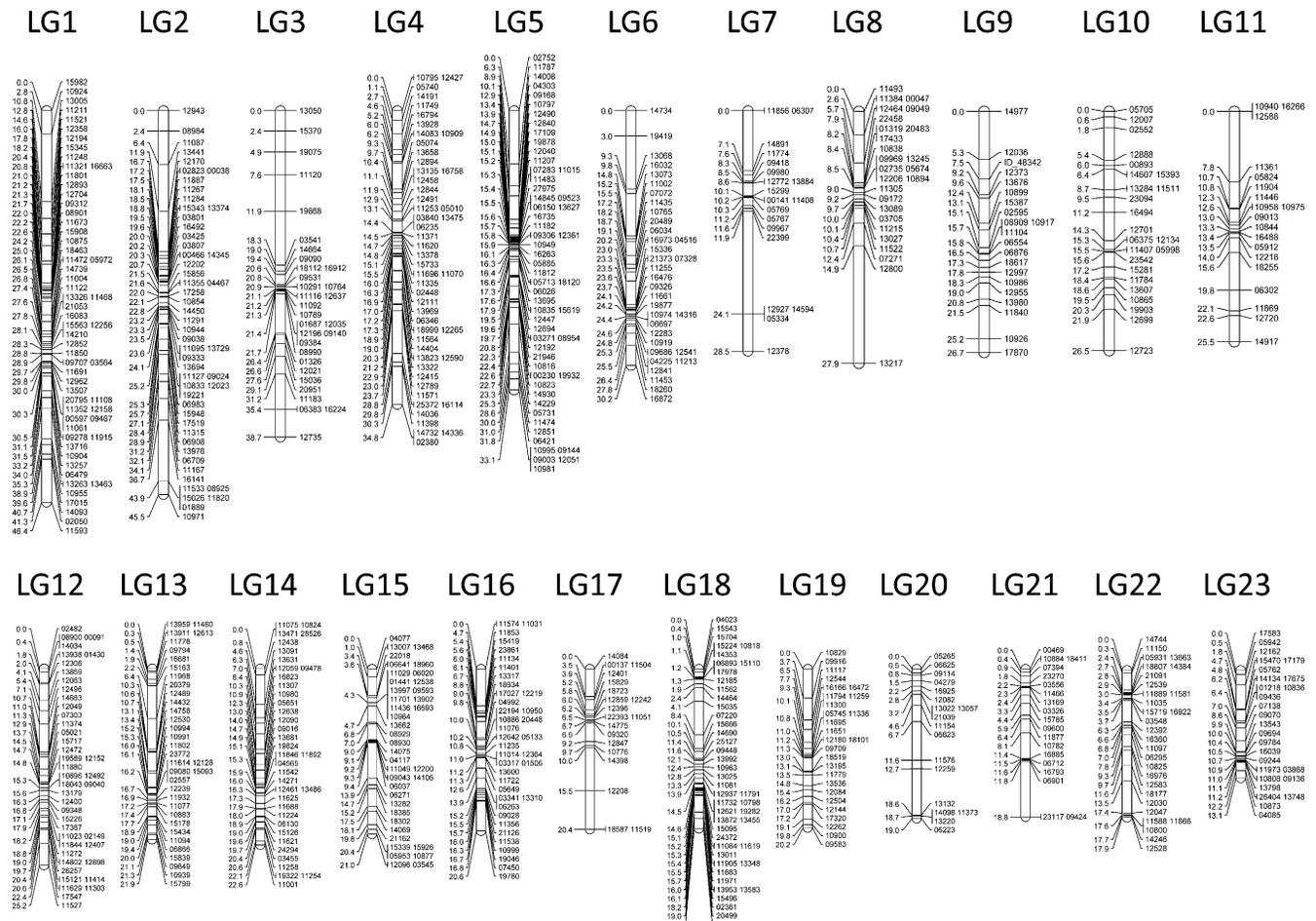


Figure 3 *Lucania parva* linkage map. Numbers on the right of each linkage group indicate the marker name and numbers on the left indicate the position in centimorgans (cM).

the two *L. goodei* populations, so we also genotyped the F<sub>0</sub> grandparents to clarify phases.

### Linkage map creation

For each species, individual maps were constructed for each individual F<sub>2</sub> family (14 for *L. goodei*; 8 for *L. parva*) using JoinMap 4.1 (Li *et al.* 2008). Parental and grandparental genotypes provided phase information for each locus. Grouping thresholds of LOD 4.0 (*L. goodei*) and 3.5 (*L. parva*) were used and markers significantly out of Hardy-Weinberg equilibrium ( $P < 0.001$ ) were excluded. The Kosambi mapping function (Kosambi 1944) was used to convert recombination frequencies to cM. For each species, a consensus map was constructed using the Map Integration tool in JoinMap from the individual family map inputs. Linkage groups from individual families were joined if they shared two or more markers. MapChart 2.2 (Stam 1993) was used for graphical representation of the consensus map for each species. *Lucania parva* LGs were numbered in descending order based on total length in cM. *Lucania goodei* LGs were numbered based upon synteny with *L. parva* LGs (see section: *Synteny comparisons*).

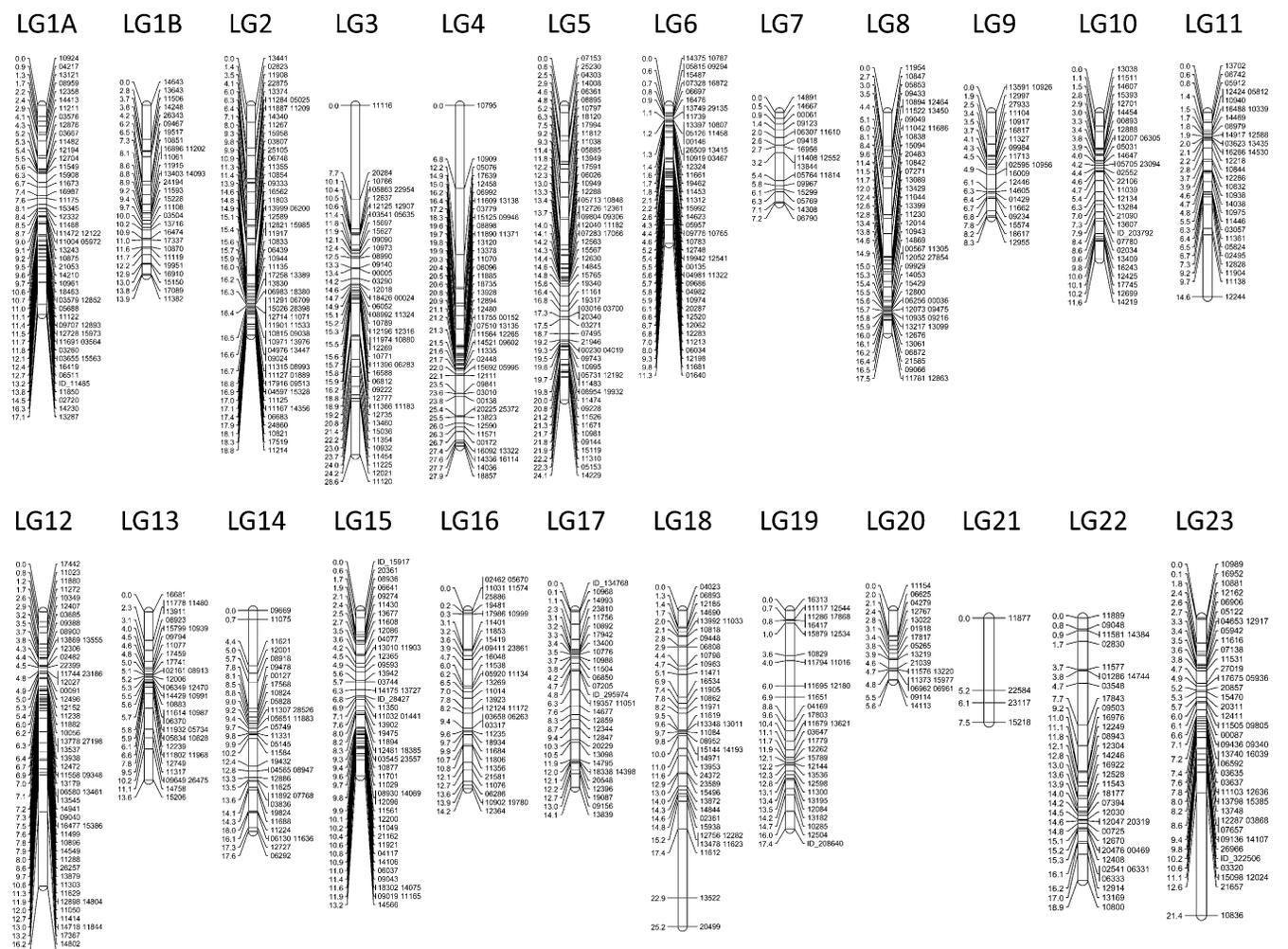
### Synteny comparisons

To compare synteny between *L. parva* and *L. goodei*, SNPs designed from a common contig were considered to be putatively orthologous.

Consistent clusters of SNPs from common contigs in both species provided evidence of synteny at the LG level (*i.e.*, SNPs from the same contigs clustered together in both species). Within LGs, marker order was compared among orthologous SNPs by performing rank order correlations between species. *Lucania goodei* and *L. parva* LGs also were compared with chromosomes from medaka (*Oryzias latipes*), the most closely related species with a fully sequenced genome [both are in the superorder Acanthopterygii (Steinke *et al.* 2006)]. Sequences of contigs in the *Lucania* linkage maps were blasted against medaka sequences using Blast. If a given contig had highly significant blast hits (bit score > 100) against a single medaka LG, then its approximate position was estimated in the medaka genome. However, if a given contig had multiple, highly significant blast hits on multiple medaka LGs, then the orthologous location in the medaka genome could not be determined. Hence, synteny was examined at the level of LG clustering for the *L. goodei* and *L. parva*, *L. goodei* and medaka, and *L. parva* and medaka comparisons. Synteny was examined at the level marker order for only the *L. goodei* and *L. parva* comparison. False-discovery rate *P*-values were calculated for rank order correlations using *fdrtool* in R (Strimmer 2008).

### Flow cytometry

The genome size of both species was estimated using flow cytometry to determine how much recombination was occurring per megabase.



**Figure 4** *Lucania goodei* linkage map. Numbers on the right of each linkage group indicate the marker name and numbers on the left indicate the position in centimorgans (cM).

Four *L. goodei* and four *L. parva* individuals were collected from the Lower Bridge site on the Wakulla River, Florida. The DNA content of erythrocyte nuclei was measured using flow cytometry at Ursinus College (see File S1). *Betta splendens* was used as a standard.

## RESULTS

The two *Lucania* species differed in chromosome number. *Lucania parva* had 23 chromosomes (Figure 2A). Twenty-two of these were acrocentric chromosomes of similar size, and one was a metacentric chromosome approximately twice the size of the other chromosomes (denoted by an arrow in Figure 2A). *Lucania goodei* had 24 acrocentric chromosomes (Figure 2B).

The number of chromosomes corresponded well to the number of LGs that were recovered. For *L. parva*, 23 linkage groups were found, which matched the number of chromosomes observed in the karyotype. Specifically, 766 SNP markers were resolved into 23 LGs (Table 1; Figure 3). Many of these SNP markers came from an EST that corresponded to a known protein, many of which could be assigned a putative function (Table S1). The number of markers per linkage group (LG) ranged from 18 (LG 20) to 59 (LG 1). The total length of the map was 605 cM with the average LG being 26.3 cM (Table 1). Marker density was 1.26 markers per cM on average spanning from 0.67 (LG 7) to 2.22 (LG 18). Marker density was relatively consistent with the largest gap being 13 cM on LG 8 and with 32 gaps of 4 cM or larger across the map. Average genome size (C-value) in *L. parva* was 1.423 pg (range: 1.396-1.450). One centimorgan in *L. parva* is thus approximately 2.3 Mb.

Similarly, the number of LGs recovered for *L. goodei* matched the number of chromosomes (1N = 24). Significant linkages were found for 915 SNP markers making up 24 LGs (Table 1; Figure 4). Again, many of these SNP markers came from an EST that corresponded to a known protein (Table S1). The number of markers per LG ranged from 4 (LG 21) to 66 (LG 2). The linkage map spanned 392 cM with the average LG size being 16.33 cM (Table 1). Average marker density was 2.41 markers per cM, ranging from 0.53 (LG 21) to 4.42 (LG 6).

Marker density was consistent with the largest gap being 8.7 cM on LG 23 and only 9 gaps of 4 cM or larger across the entire map. Average genome size in *L. goodei* was 1.349 pg (range: 1.342-1.356). One centimorgan in *L. goodei* is thus approximately 3.44 Mb.

## Synteny of *L. parva* and *L. goodei* maps

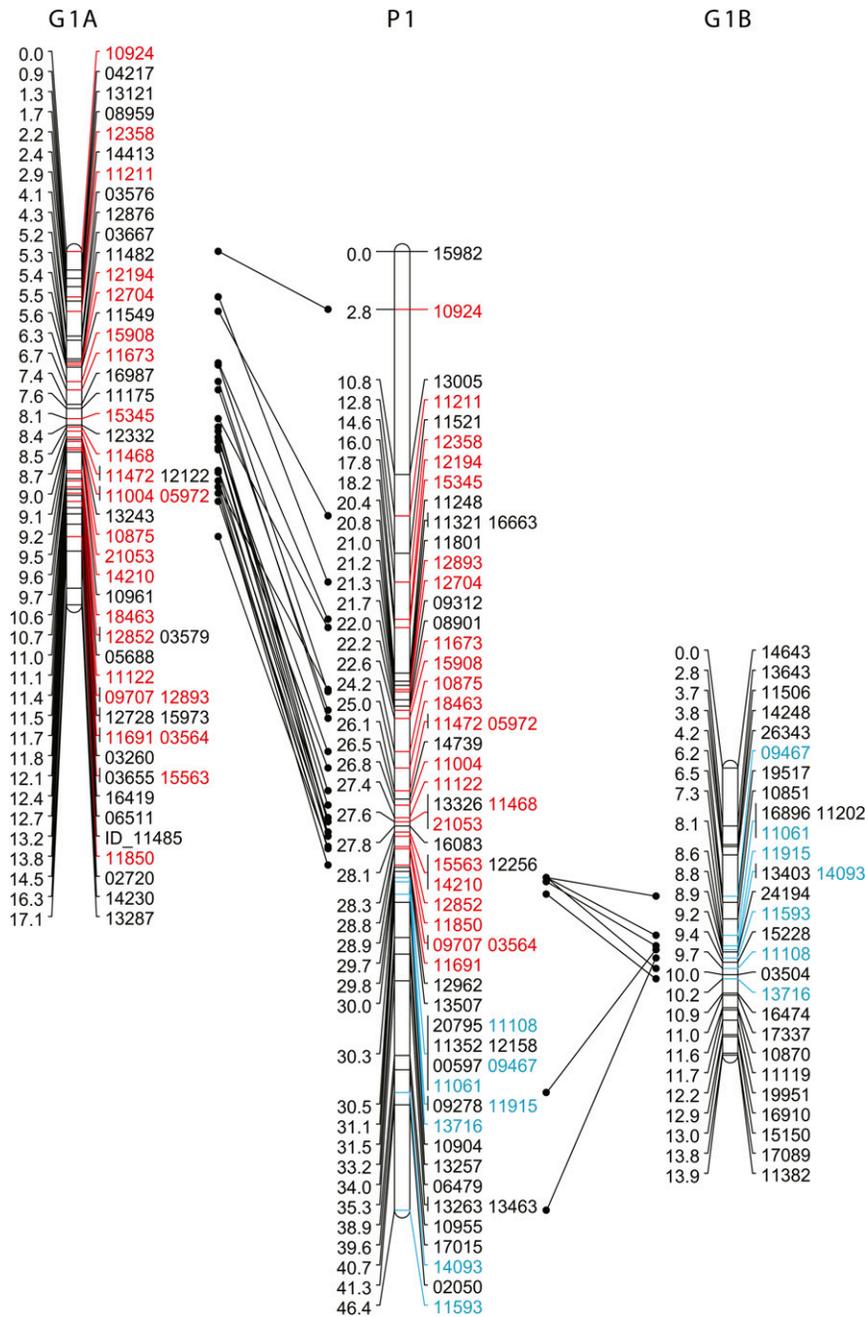
The LGs were highly syntenic between *L. goodei* and *L. parva*, allowing us to unambiguously assign orthologous LGs. Across all LGs, we found 368 markers shared between the linkage maps that were from putatively orthologous ESTs. Figure 5 shows that markers from the putatively orthologous ESTs clustered together in the same LGs in *L. goodei* and *L. parva*. Specifically, 364 (98.91%) markers (those from a common contig and in both linkage maps) showed a pattern of synteny, whereas only 4 markers (1.09%) deviated and clustered differently in the two species (Figure 5 and Figure S1).

Two LGs from *L. goodei* were syntenic with the largest LG in *L. parva*, strongly suggesting that LG 1 represents the metacentric chromosome observed in the *L. parva* metaphase spread. Figure 6 shows that LG 1A in *L. goodei* was syntenic with the top portion of LG 1 in *L. parva* (matching at 24 markers: Figure 5), and LG 1B was syntenic with the bottom portion (matching at 7 markers).

Synteny was less conserved at the level of marker order within LGs (Table 2). Rank order correlations were high for some LGs but were low and nonsignificant for others. Across all LGs combined, marker order was highly correlated in *L. parva* and *L. goodei* (Spearman rank order correlation:  $n = 364$ ,  $r = 0.99$ ,  $P < 0.0001$ ). When rank order correlations were performed on LGs separately (and corrected for multiple testing by use of the false-discovery rate), marker order was significantly correlated in 11 of 23 syntenic LGs (50%; Table 2; note that LG21 lacked sufficient orthologous markers to test marker order). The marker order correlations were not statistically significantly greater than zero for the other 12 LGs. To determine whether noncoding repetitive RNAs or slight differences in the order of densely mapped markers were influencing this result, we repeated these synteny analyses using only syntenic markers  $\geq 0.5$  cM apart in either species that

		<i>L. goodei</i> LG																								
		1A	1B	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
<i>L. parva</i> LG																										
1		24	7																							
2				27																						
3					13																					
4						22																				
5							28																			
6								14																		
7									7							1										
8										11																
9											7															
10												14														
11													13													
12														25												
13															17											
14																14										
15																	22									
16																		18								
17																			7							
18																					22					
19																						15				
20																							10			
21																								2	3	
22																									13	
23																										12
medaka #		3	11	9	5	8	16	4	20	17	15	12	21	7	6	24	14	10	13	19	23	18	2	22	1	

**Figure 5** Summary of synteny comparisons between *L. parva* and *L. goodei* linkage groups. Bolded numbers along the diagonal show the number of orthologous single-nucleotide polymorphisms on the linkage groups. Numbers off the diagonal are nonsyntenic markers. Identity of syntenic medaka chromosomes listed along the bottom row.



**Figure 6** Synteny of linkage group 1 in *L. parva* and 1A,B in *L. goodei*. Orthologous single-nucleotide polymorphisms (SNPs) between *L. parva* linkage group 1 (P: middle) and *L. goodei* (G) linkage groups 1A (far left) and 1B (far right) are shown connected by lines. SNPs syntenic with *L. goodei* 1A are highlighted in red; SNPs syntenic with *L. goodei* 1B are highlighted in blue.

had a single identified location in the medaka genome. Again, we found significant marker order preservation in less than half (38%) of the groups (6 of 16 with more than 5 syntenic markers).

### Synteny of *Lucania* and medaka

Synteny at the LG level was also well preserved between *Lucania* and medaka (Figure S2). Medaka has 24 chromosomes, and all *Lucania* LGs had a single best match to a medaka chromosome (listed at the bottom of Figure 5). Between *L. goodei* and medaka 559 of 585 (95.6%) orthologous markers were syntenic at the LG level. Similarly, *L. parva* had 532 of 545 (97.6%) orthologous markers that were found to be syntenic with medaka chromosomes. LG 1A in *L. goodei* corresponded to chromosome 3 in medaka and LG 1B corresponded to medaka chromosome 11.

### DISCUSSION

Using high-throughput Illumina Infinium genotyping assays, we created two SNP-based linkage maps for *Lucania parva* and *L. goodei* with high marker density (1.26 markers/cM in *L. parva* and 2.41 markers/cM in *L. goodei*). These linkage maps establish genomic resources for *Lucania* and provide the groundwork for future linkage disequilibrium studies, quantitative trait loci mapping, molecular population genetic studies, and further synteny comparisons with other teleost species. The fact that many of these SNPs came from ESTs whose protein functions are known in other groups allows us to estimate the position of functionally important loci in *Lucania* (see Table S1). This is the first linkage map for any member of the Fundulidae family, a group that exhibits an extraordinary ability to tolerate and adapt to physiological extremes (Burnett *et al.* 2007).

■ Table 2 Marker order correlations for syntenic markers on each linkage group in *Lucania parva* and *L. goodei*

LG	Size <i>L. parva</i> , cM	Size <i>L. goodei</i> , cM	No. Syntenic Markers	Spearman r	P Value	FDR-Corrected P Value
1	46.4	A:17.1	24	0.8183	0.000001	0.000084
		B:13.9	7	0.607	0.1362	0.251906
2	45.5	22	27	0.835	0.000001	0.000084
3	38.7	28.6	13	0.4341	0.137904	0.251906
4	34.8	27.9	22	0.9198	0.000001	0.000084
5	33.5	26	28	0.7225	0.000014	0.007719
6	30.2	11.3	14	0.033	0.914228	1.000000
7	28.5	7.2	7	0.785714	0.0536	0.251906
8	27.9	17.5	11	0.6818	0.02071	0.030374
9	26.7	8.3	7	0.214286	0.5962	1.000000
10	26.5	11.6	14	0.3626	0.20193	1.000000
11	25.5	14.6	13	0.1923	0.529033	1.000000
12	25.2	22.2	25	0.5354	0.005817	0.030374
13	22.7	13.8	17	0.348	0.170412	0.251906
14	22.6	18.4	14	0.5956	0.02454	0.094289
15	21	13.2	21	0.441	0.039729	0.094289
16	20.6	15.3	18	0.5542	0.017116	0.030374
17	20.4	14.1	7	0.142857	0.7264	1.000000
18	20.3	30	22	0.8701	0.000001	0.000084
19	20.2	17.9	15	0.7214	0.002382	0.007719
20	19	9.5	10	0.3455	0.328748	1.000000
21	18.8	7.5	2	N/A	N/A	N/A
22	17.9	22.1	13	0.7637	0.002393	0.021931
23	13.1	21.4	12	0.6783	0.0153	0.030374

LG, linkage group; FDR, false-discovery rate.

By combining our linkage maps with actual chromosome counts, we found strong evidence for a major chromosomal rearrangement between *L. parva* and *L. goodei*. Our metaphase spread showed that a large metacentric chromosome was present in *L. parva* and absent in *L. goodei* (see arrow in Figure 2). Our linkage maps indicate that the largest *L. parva* LG (LG 1, 46.4 cM) is syntenic with two smaller *L. goodei* chromosomes (LG 1A and 1B). Comparisons of interspecific linkage maps have previously been used in other taxa to identify chromosomal rearrangements between species (Tanksley *et al.* 1992; Burke *et al.* 2004). Several pieces of evidence suggest that the metacentric chromosome is due to a Robertsonian fusion of two smaller acrocentric chromosomes in the *L. parva* lineage rather than a chromosomal fission event in the *L. goodei* lineage (a metacentric chromosome becoming two acrocentric chromosomes). First, *Fundulus parvipinnis* is the closest relative to *Lucania* (Whitehead 2010), and its karyotype is similar to *L. goodei* with 1N = 24 (Chen 1971). Second, *L. goodei* LG 1A is syntenic to chromosome 3 in medaka, and LG 1B is syntenic to chromosome 11. The fact that each of these two LGs map to a single LG in medaka suggests that they are unlikely to reflect the outcome of a fission event. Our evidence for a chromosomal fusion clarifies a previous report (made without any supporting data) that *L. parva*'s karyotype of 1N = 23 deviates from the typical fundulid karyotype of 1N = 24 (Uyeno and Miller 1971). However, our evidence for a fusion should be further verified using *in situ* hybridization to determine that LG1 markers are indeed found on the metacentric chromosome in *L. parva*.

With the exception of the fused chromosome, comparisons between our maps reveal that large-scale structure is widely conserved between species. At the level of the linkage group, we found high preservation of synteny between *Lucania* linkage groups. We also found preservation of synteny between *Lucania* and medaka chromosomes. Each *Lucania* chromosome could be assigned unambiguously as syntenic to a single medaka chromosome. This result is consistent with findings from other teleost linkage maps and genomes (Schartl

*et al.* 2013). This degree of synteny will facilitate future comparisons with other related teleost species for which genomic resources are emerging, such as guppies (Tripathi *et al.* 2009; Fraser *et al.* 2011) and fundulids (Cossins and Crawford 2005; Burnett *et al.* 2007). Currently, the fusion we document in *Lucania* (between chromosomes that are syntenic to medaka 3 and 11) is unique among related fish species for which linkage maps exist. Guppies (*Poecilia reticulata*) also possess a fused chromosome, but it has occurred between two chromosomes that are syntenic to medaka 2 and 21 (Tripathi *et al.* 2009). Tilapia (*Oreochromis niloticus*) possess two fused chromosomes: one between chromosomes that are syntenic to medaka 2 and 4, and another between chromosomes syntenic to medaka 6 and 12 (Liu *et al.* 2013).

Our comparisons of synteny at the marker order level revealed significant marker order preservation in only half of the linkage groups between *L. parva* and *L. goodei*. Our estimate of 50% collinear markers is much lower than a previous estimate of marker order preservation between hybridizing populations, which found that 83% of markers between incipient species of whitefish (*Coregonus clupeaformis*) were collinear (Rogers *et al.* 2007). Comparisons of chinook salmon and rainbow trout (genus *Oncorhynchus*) also suggest high preservation of marker order between congeners (Naish *et al.* 2013). Our study differs from these previous ones because we used SNPs derived from ESTs rather than microsatellites and consequently had a much denser distribution of markers on our map. However, further work is needed to determine if the low marker-order synteny in *Lucania* is genuine or simply the result of (1) low number of orthologous markers on some linkage groups, (2) genotyping errors, or (3) map construction errors due to merging maps from multiple families within each species. If genuine, then the low marker-order synteny found in *Lucania* would be indicative of small-scale rearrangements, which could potentially contribute to the reduction of gene flow and the evolution of reproductive isolation.

Our linkage maps for *Lucania goodei* and *L. parva* showed that a large-scale genomic rearrangement has occurred between species. This Robertsonian fusion may have aided divergence in these closely related species and helped to maintain species boundaries in zones of contact by suppressing recombination in hybrids. Other fundulid species also differ in karyotype, and our work may give insight into the mechanisms by which these changes occur. These maps will enable the use of numerous genomic techniques to determine how reproductive isolation has evolved in *Lucania*. The maps will also further a general understanding of the evolution of many other unique traits in this group including vision, color pattern, and salinity tolerance.

## ACKNOWLEDGMENTS

Leslie Noa and Jessie Coonley assisted with F<sub>2</sub> crosses, fish rearing, and preservation. Matthew Schrader and Leslie Noa assisted with RNA extractions. Mark Band, Tatsiana Akraiko (Infinium assay), and Alvaro Hernandez (library preparation and sequencing) and other staff at the Keck Center for Comparative and Functional Genomics at the University of Illinois performed the sequencing and genotyping. James Walsh assisted with bioinformatics and SNP calling; Lei-ting Li helped with Joinmap. Sairaghav Nissankula assisted with blast searches. James Urton and Shaun McCann provided invaluable advice and help with the karyotype. Katie Peichel provided helpful comments on the manuscript. This work was approved by the University of Illinois IACUC (Protocol Nos. 08183, 09306, 11143). This work was funded by the University of Illinois (Research Board Award 11153) and the National Science Foundation (IOB 0445127, DEB 0953716, DEB 1110658). The 454 *L. goodei* reference assembly and the joinmap files are accessible on Dryad with this paper (doi: 10.5061/dryad.80822).

## LITERATURE CITED

Backstrom, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer *et al.*, 2008 A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics* 179: 1479–1495.

Berdan, E. L., and R. C. Fuller, 2012 Interspecific divergence of ionoregulatory physiology in killifish: insight into adaptation and speciation. *J. Zool.* 287: 283–291.

Burke, J. M., Z. Lai, M. Salmaso, T. Nakazato, S. X. Tang *et al.*, 2004 Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167: 449–457.

Burnett, K. G., L. J. Bain, W. S. Baldwin, G. V. Callard, S. Cohen *et al.*, 2007 *Fundulus* as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 2: 257–286.

Butlin, R. K., 2005 Recombination and speciation. *Mol. Ecol.* 14: 2621–2635.

Carleton, K. L., 2011 Quantification of transcript levels with quantitative RT-PCR, pp. 279–295 in *Molecular Methods for Evolutionary Genetics*, edited by V. Orgogozo, and M. V. Rockman. Springer, New York.

Chen, T. R., 1971 A comparative chromosome study of twenty killifish species of the genus *Fundulus* (Teleostei: Cyprinodontidae). *Chromosoma* 32: 436–453.

Coluzzi, M., A. Sabatini, A. della Torre, M. A. Di Deco, and V. Petrarca, 2002 A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298: 1415–1418.

Cossins, A. R., and D. L. Crawford, 2005 Opinion - Fish as models for environmental genomics. *Nat. Rev. Genet.* 6: 324–333.

Duggins, C. F., A. A. Karlin, and K. G. Relyea, 1983 Electrophoretic variation in the killifish genus *Lucania*. *Copeia* 564–570.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Pearson Education Limited, New York, NY.

Faria, R., and A. Navarro, 2010 Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* 25: 660–669.

Feder, J. L., F. B. Roethele, K. Filchak, J. Niedbalski, and J. Romero-Severson, 2003 Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163: 939–953.

Foley, B. R., C. G. Rose, D. E. Rundle, W. Leong, G. W. Moy *et al.*, 2011 A gene-based SNP resource and linkage map for the copepod *Tigriopus californicus*. *BMC Genomics* 12: 12.

Fraser, B. A., C. J. Weadick, I. Janowitz, F. H. Rodd, and K. A. Hughes, 2011 Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genomics* 12: 568.

Fuller, R. C., 2008 Genetic incompatibilities in killifish and the role of environment. *Evolution* 62: 3056–3068.

Fuller, R. C., and L. A. Noa, 2008 Distribution and stability of sympatric populations of *Lucania goodei* and *L. parva* across Florida. *Copeia* 699–707.

Fuller, R. C., and K. M. Claricoates, 2011 Rapid light-induced shifts in opsin expression: finding new opsins, discerning mechanisms of change, and implications for visual sensitivity. *Mol. Ecol.* 20: 3321–3335.

Fuller, R. C., K. E. McGhee, and M. Schrader, 2007 Speciation in killifish and the role of salt tolerance. *J. Evol. Biol.* 20: 1962–1975.

Fuller, R. C., L. A. Noa, and R. S. Strellner, 2010 Teasing apart the many effects of lighting environment on opsin expression and foraging preference in bluefin killifish. *Am. Nat.* 176: 1–13.

Gale, M. D., and K. M. Devos, 1998 Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* 95: 1971–1974.

Hubbs, C. L., B. W. Walker, and R. E. Johnson, 1943 Hybridization in nature between species of American cyprinodont fishes. *Contr. Lab. Vertebrate Biol.* 23: 1–21.

Jaari, S., M. H. Li, and J. Merila, 2009 A first-generation microsatellite-based genetic linkage map of the Siberian jay (*Perisoreus infaustus*): insights into avian genome evolution. *BMC Genomics* 10: 1.

Kasahara, M., K. Naruse, S. Sasaki, Y. Nakatani, W. Qu *et al.*, 2007 The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447: 714–719.

Kitano, J., J. A. Ross, S. Mori, M. Kume, F. C. Jones *et al.*, 2009 A role for a neo-sex chromosome in stickleback speciation. *Nature* 461: 1079–1083.

Kosambi, D. D., 1944 The estimation of map distance from recombination values. *Ann. Eugen.* 12: 172–175.

Kozak, G. M., A. B. Rudolph, B. L. Colon, and R. C. Fuller, 2012 Postzygotic isolation evolves before prezygotic isolation between fresh and saltwater populations of the rainwater killifish, *Lucania parva*. *Int. J. Evol. Biol.* 2012: 523967.

Kozak, G. M., R. S. Brennan, E. L. Berdan, R. C. Fuller, and A. Whitehead, 2014 Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution* 68: 63–80.

Lee, D. S., C. R. Gilbert, C. H. Hocutt, R. E. Jenkins, D. E. McAllister *et al.*, 1980 *Atlas of North American Freshwater Fishes*. North Carolina State Museum of Natural History, Raleigh, NC.

Lee, S. F., L. Rako, and A. A. Hoffmann, 2011 Genetic mapping of adaptive wing size variation in *Drosophila simulans*. *Heredity* 107: 22–29.

Li, H., J. Ruan, and R. Durbin, 2008 Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18: 1851–1858.

Liu, F., F. Sun, J. Li, J. H. Xia, G. Lin *et al.*, 2013 A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis* spp.) and mapping of sex-determining loci. *BMC Genomics* 14: 58.

Lubieniecki, K. P., S. L. Jones, E. A. Davidson, J. Park, B. F. Koop *et al.*, 2010 Comparative genomic analysis of Atlantic salmon, *Salmo salar*, from Europe and North America. *BMC Genet.* 11: 105.

Lucas, M. R., N. N. Diop, S. Wanamaker, J. D. Ehlers, P. A. Roberts *et al.*, 2011 Cowpea-Soybean synteny clarified through an improved genetic map. *Plant Genome* 4: 218–225.

McGraw, L. A., J. K. Davis, L. J. Young, and J. W. Thomas, 2011 A genetic linkage map and comparative mapping of the prairie vole (*Microtus ochrogaster*) genome. *BMC Genet.* 12: 60.

Ming, R., S. C. Liu, Y. R. Lin, J. da Silva, W. Wilson *et al.*, 1998 Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. *Genetics* 150: 1663–1682.

- Muchero, W., N. N. Diop, P. R. Bhat, R. D. Fenton, S. Wanamaker *et al.*, 2009 A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proc. Natl. Acad. Sci. USA* 106: 18159–18164.
- Naish, K. A., R. B. Phillips, M. S. O. Briec, L. R. Newton, A. E. Elz *et al.*, 2013 Comparative genome mapping between chinook salmon (*Oncorhynchus tshawytscha*) and rainbow trout (*O. mykiss*) based on homologous microsatellite loci. *G3 (Bethesda)* 3: 2281–2288.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland, 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* 98: 12084–12088.
- Peichel, C. L., K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. *Nature* 414: 901–905.
- Prince, J. P., E. Pochard, and S. D. Tanksley, 1993 Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. *Genome* 36: 404–417.
- Rieseberg, L. H., 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16: 351–358.
- Rieseberg, L. H., J. Whitton, and K. Gardner, 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713–727.
- Rogers, S. M., N. Isabel, and L. Bernatchez, 2007 Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics* 175: 375–398.
- Schartl, M., R. B. Walter, Y. J. Shen, T. Garcia, J. Catchen *et al.*, 2013 The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* 45: 567–572.
- Stam, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package - Joinmap. *Plant J.* 3: 739–744.
- Stapley, J., T. R. Birkhead, T. Burke, and J. Slate, 2008 A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics* 179: 651–667.
- Steinke, D., W. Salzburger, and A. Meyer, 2006 Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J. Mol. Evol.* 62: 772–784.
- Strimmer, K., 2008 *fdrtool*: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* 24: 1461–1462.
- Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141–1160.
- Timusk, E. R., M. M. Ferguson, H. K. Moghadam, J. D. Norman, C. C. Wilson *et al.*, 2011 Genome evolution in the fish family salmonidae: generation of a brook charr genetic map and comparisons among charrs (Arctic charr and brook charr) with rainbow trout. *BMC Genet.* 12: 68.
- Tripathi, N., M. Hoffmann, E. M. Willing, C. Lanz, D. Weigel *et al.*, 2009 Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proc. Biol. Sci.* 276: 2195–2208.
- Uyeno, T., and R. R. Miller, 1971 Multiple sex chromosomes in a Mexican cyprinodontid fish. *Nature* 231: 452–453.
- Wang, J., B. Roe, S. Macmil, Q. Yu, J. E. Murray *et al.*, 2010 Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11: 261.
- Whitehead, A., 2010 The evolutionary radiation of diverse osmotolerant physiologies in killifish (*Fundulus* Sp.). *Evolution* 64: 2070–2085.

Communicating editor: A. S. McCallion