Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication

Angel Amores,* Julian Catchen,† Allyse Ferrara,‡ Quenton Fontenot,‡ and John H. Postlethwait*,[1]
*Institute of Neuroscience and †Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, and ‡Nicholls State University, Department of Biological Sciences, Thibodaux, Louisiana 70310

ABSTRACT Genomic resources for hundreds of species of evolutionary, agricultural, economic, and medical importance are unavailable due to the expense of well-assembled genome sequences and difficulties with multigenerational studies. Teleost fish provide many models for human disease but possess anciently duplicated genomes that sometimes obfuscate connectivity. Genomic information representing a fish lineage that diverged before the teleost genome duplication (TGD) would provide an outgroup for exploring the mechanisms of evolution after whole-genome duplication. We exploited massively parallel DNA sequencing to develop meiotic maps with thrift and speed by genotyping F_1 offspring of a single female and a single male spotted gar (*Lepisosteus oculatus*) collected directly from nature utilizing only polymorphisms existing in these two wild individuals. Using Stacks, software that automates the calling of genotypes from polymorphisms assayed by Illumina sequencing, we constructed a map containing 8406 markers. RNA-seq on two map-cross larvae provided a reference transcriptome that identified nearly 1000 mapped protein-coding markers and allowed genome-wide analysis of conserved synteny. Results showed that the gar lineage diverged from teleosts before the TGD and its genome is organized more similarly to that of humans than teleosts. Thus, spotted gar provides a critical link between medical models in teleost fish, to which gar is biologically similar, and humans, to which gar is genomically similar. Application of our F_1 dense mapping strategy to species with no prior genome information promises to facilitate comparative genomics and provide a scaffold for ordering the numerous contigs arising from next generation genome sequencing.

TELEOST fish provide numerous medical models. Some are induced mutant models, as in zebrafish and medaka (*i.e.*, Moore *et al.* 2006; Schartl *et al.* 2010). Others are evolutionary mutant models, in which naturally occurring mutations lead to adaptive phenotypes that mimic human disease (Albertson *et al.* 2009), as in cichlids (craniofacial anomalies), platyfish (melanoma), mollies (premature puberty), cavefish (retinal degeneration), and icefish (osteopenia and anemia) (Eastman and Devries 1981; Streelman *et al.* 2003; Meierjohann and Schartl 2006; Near *et al.* 2006; Jeffery 2009; Valenzano *et al.* 2009; Albertson *et al.* 2010; Lampert *et al.* 2010; Zhang *et al.* 2010b). Teleost genomes differ from mammalian genomes, however, by a whole-genome duplication event, the teleost genome duplication (TGD) (Figure 1) (Amores *et al.* 1998; Postlethwait *et al.* 1998; Taylor *et al.* 2003; Jaillon *et al.* 2004). While the TGD can facilitate the dissection of ancestral gene functions due to the partitioning of ancestral subfunctions in the course of evolution (Force *et al.* 1999; Postlethwait *et al.* 2004), it can also obfuscate correlations between teleost disease models and their human counterparts because of the difficulty of ortholog assignment after lineage-specific loss of duplicated genes and the asymmetric evolution of gene duplicates. Genomic resources from a ray-fin (Actinopterygian) fish that diverged from teleosts before the TGD (Figure 1) would facilitate the connectivity of teleost and mammalian genomes. Unfortunately, candidate species for this role, including polypterus, paddlefish, sturgeons, bowfin, and gar (Blacklidge and Bidwell 1993; Inoue *et al.* 2003;

doi: 10.1534/genetics.111.127324
Manuscript received January 29, 2011; accepted for publication May 16, 2011
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.111.127324/-/DC1.
Data in public repositories: SRA026509.1
IRB (IACUC) number: 08-13RA
[1]Corresponding author: Institute of Neuroscience, University of Oregon, Eugene, OR 97403. E-mail: jpostle@uoregon.edu

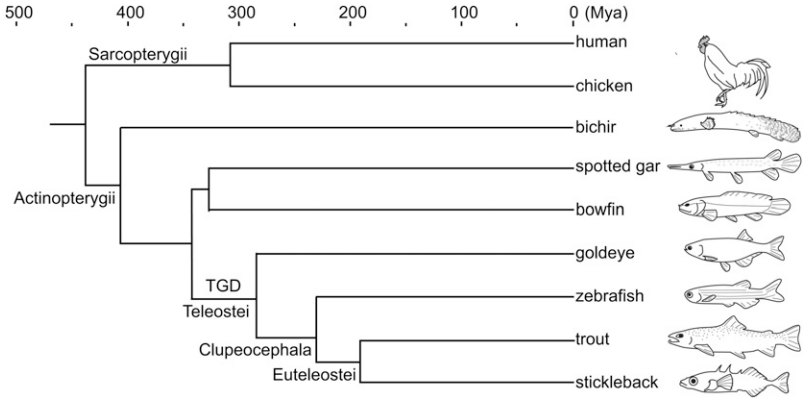

Figure 1 Phylogenetic relationships among lobe-fin (Sarcopterygii) and ray-fin (Actinopterygii) fish (after Inoue *et al.* 2005; Benton and Donoghue 2007). Lineages of teleosts and gar diverged ~340 million years ago (MYA) and lineages of ray-fin fish and tetrapods separated ~440 MYA. Species: human (*Homo sapiens*), chicken (*Gallus gallus*), bichir (*Polypterus ornatipinnis*), spotted gar (*Lepisosteus oculatus*), bowfin (*Amia calva*), goldeye (*Hiodon alosoides*), zebrafish (*Danio rerio*), trout (*Oncorhynchus mykiss*), and stickleback (*Gasterosteus aculatus*).

Hardie and Hebert 2004) have virtually no genome resources and have life history traits inconvenient for the construction of large-scale genetic maps.

Teleost genomes possess substantially rearranged chromosomes with respect to mammalian chromosomes (Postlethwait *et al.* 1998; Nakatani *et al.* 2007), and although it has been suggested that chromosome rearrangements accelerated after the TGD, this idea is controversial (Comai 2005; Semon and Wolfe 2007; Hufton *et al.* 2008). Comparative analysis of a fish genome occupying a lineage that diverged from teleosts shortly before the TGD would test whether chromosome rearrangements detected in teleosts arose before or after the TGD.

Analysis of a half dozen genes in spotted gar (*Lepisosteus oculatus*), a large, air-breathing ray-finned North American fish, suggested that its lineage might have diverged from the teleost lineage before the TGD (Hoegg *et al.* 2004; Crow *et al.* 2006). If gar did diverge before the TGD, then gar would provide a genomic intermediary between teleost medical models and the human genome. Simple hormone injections can induce wild-caught gar to spawn in the laboratory and each female can produce on average more than 6000 fertilized eggs (Smith 2006). Gar embryos are suitable for *in situ* hybridization studies and develop in the lab to hatching and beyond. Spotted gar, however, like many plant and animal species of ecological, evolutionary, agricultural, pharmacological, behavioral, and medical interest, has inconvenient life history traits (males mature when 1 yr old and females when 2 yr old; Smith 2006) and gar lacks genome resources (just six nuclear gene sequences in GenBank). Furthermore, current meiotic mapping methods produce only a few hundred markers at great expense, usually do not provide protein-coding loci, and require multigenerational pedigrees (Kucuktas *et al.* 2009; Sanetra *et al.* 2009; Tripathi *et al.* 2009; Chintamanani *et al.* 2010; Li *et al.* 2010; Okada *et al.* 2010). While next generation sequencing methodologies have been effectively used for resequencing genomes of canonized model organisms with sequenced genomes (Hobert 2010; Zuryn *et al.* 2010) and for analyzing populations in genome-wide association studies (Hohenlohe *et al.* 2010), the short sequences these methods generate have been thought to be less suitable for species that lack a reference genome.

To address these problems, we devised novel strategies that utilize next-generation sequencing and Stacks software, which converts short-read sequences into called genotypes (Catchen *et al.* 2011), to economically create genetic maps containing many thousands of mapped markers. Because the gar's 1- to 2-yr generation time erects a disincentive to the construction of a traditional F_2 or backcross mapping panel, we capitalized on polymorphisms naturally present in the genome of a single male and a single female spotted gar taken directly from Louisiana bayous to develop a genetic map directly by genotyping their F_1 progeny. We used as markers polymorphisms that were present in restriction-associated DNA (RAD)-tag sequences adjacent to restriction enzyme cut sites (Miller *et al.* 2007a,b; Baird *et al.* 2008). We associated mapped RAD-tag markers to gar coding genes by constructing a reference transcriptome from one embryo and the head of one larva. These datasets allowed the identification of nearly 1000 mapped markers as representatives of protein-coding genes. We used coding markers on the gar map to test the hypothesis that the gar lineage diverged from the teleost lineage before the TGD and to challenge the hypothesis that most rearrangements in teleost genomes occurred after the teleost genome duplication.

The mapping strategy developed here is directly applicable to numerous nonmodel organisms and their application should spur genomic research on previously intractable species. Furthermore, the great number of genetic markers mapped can help order the thousands of unordered contigs that arise in next generation genome sequencing projects.

Materials and Methods

Animals

A single female and a single male adult spotted gar (*L. oculatus*) were collected in Louisiana (Bayous Chevreuil and Gheens, respectively) and maintained in a recirculating system at 23–25° on a 14-h-light/10-h-dark photoperiod. Injecting fish with Ovaprim at 0.5 ml/kg body weight induced spawning within 48 hr. This mating produced thousands of progeny, of which we collected 500 for DNA extractions and selected 94 of these to genotype for our F_1 map cross and two others to use for transcriptomics.

Progeny were maintained at 23–25° until 14 days postfertilization (dpf), when they were killed by MS-222 overdose and stored in 100% EtOH at −20°. Parents were killed by concussion, and brain, liver, blood, gonads, kidney, and muscle samples were collected for transcriptomics. Tissue samples were stored in 100% EtOH at −20° until isolation of genomic DNA (DNeasy Blood and Tissue Kit, Qiagen, see supporting information, File S1 for details). Local university animal care committees approved euthanasia and all other animal procedures.

RAD-tag libraries

Genomic DNA was purified from parents and progeny and digested with high-fidelity *Sbf*I (New England Biolabs), which has an 8-bp, GC-rich recognition site (CCTGCAGG) and cuts ~25,000 to 30,000 times in teleost genomes. Samples were ligated to adapters with a set of five nucleotide (nt) barcodes each different by at least two nucleotides for unambiguous assignment. RAD-tag libraries were created as described (Miller *et al.* 2007b; Baird *et al.* 2008; Hohenlohe *et al.* 2010) (see File S1 for further details). A total of 50 ng of pooled, size-selected DNA was PCR amplified for 12 cycles and the 200- to 500-bp fraction was gel purified. RAD-tag libraries were sequenced on an Illumina Genome Analyzer IIx by 80 nucleotide single end reads loading equal amounts of DNA from 10 progeny on each lane in barcoded samples.

Genotyping

Reads of low quality or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotyped using Stacks software we wrote specifically to analyze these map cross data (Catchen *et al.* 2011). Stacks is freely downloadable (http://creskolab.uoregon.edu/stacks/). The likelihood-based SNP calling algorithm (Hohenlohe *et al.* 2010) implemented in Stacks evaluates each nucleotide position in every RAD tag of all individuals, thereby differentiating true SNPs from sequencing errors. Some RAD-tag genotypes contained a single SNP, but others represented alleles that differed by multiple SNPs that were scored from these haplotypes.

Markers heterozygous in just one parent were mapped as a pseudo-testcross (Grattapaglia and Sederoff 1994) and markers heterozygous in both parents were mapped as an F_2 family. Markers segregated in four different patterns. Type $lm \times ll$ (segregating 1:1) was heterozygous in the male parent and homozygous in the female parent; $nn \times np$ (1:1) was homozygous in the male and heterozygous in the female; $hk \times hk$ (1:2:1) was heterozygous in both parents with two shared alleles; and $ef \times eg$ (1:1:1:1) was heterozygous in both parents with two sex-specific alleles and one shared allele.

Map construction

Linkage analysis was performed for markers present in at least 85 of 94 individuals (50 out of 94 for protein-coding markers) with JoinMap 4.0 (Wageningen, The Netherlands). Markers were identified as paternal or maternal, which enabled the construction of male-specific and female-specific linkage maps. Linkage between markers, recombination rate, and map distances were calculated using the Kosambi mapping function and the maximum likelihood mapping algorithm in JoinMap. Markers were grouped with an initial LOD threshold of 14.0 but many unlinked markers and small linkage groups were added using the strong crosslink feature of JoinMap at a minimum LOD of 10.0. Markers with significant segregation distortion or unlinked at LOD <10 were excluded. The graphical genotypes feature of JoinMap identified doubtful double recombinants, which were reevaluated by inspection of stacked sequences and corrected as needed; for example, some genotypes that the software called homozygotes were clearly heterozygotes with minor allele reads below threshold. Corrected genotypes were loaded into JoinMap and linkage analysis was repeated until no suspicious genotypes were identified.

The consensus map was calculated using JoinMap's population type CP (cross pollinator, or full-sib family), the Kosambi mapping function, and the regression mapping algorithm. Because JoinMap could not process >5500 markers, we selected all markers with comparative genomic information and an arbitrary set of the 8406 total linked markers to sum to 5466 markers. To reduce computational time, the largest linkage groups (1–8) were analyzed by excluding markers with identical genotypes, but smaller linkage groups were analyzed using all markers.

Transcriptomics

Total RNA was isolated from two F_1 map cross progeny that were not used for genotyping—one entire stage-30 (7 days postfertilization, near hatching) gar embryo and the head tissue of a stage-33 (12 dpf, yolk nearly exhausted) gar larva (Long and Ballard 2001)—using the RiboPure Total RNA Isolation kit (Ambion). mRNA was isolated using the MicroPolyA Purist kit (Ambion). A total of 500 ng of mRNA was reverse transcribed with SuperScript III Reverse Transcriptase (Invitrogen) and random hexamer primers (Invitrogen). Second strand cDNA was synthesized with random primers and 15 units of Klenow DNA polymerase exo-minus (Epicentre). Double stranded cDNA was sheared in a Bioruptor (Diagenode) for 30 cycles (30 sec on, 60 sec off). Sheared DNA was end repaired with the End-It DNA repair kit (Epicentre) and dA overhangs were added with Klenow DNA polymerase exo-minus. Adapters were ligated overnight and 100 ng was PCR amplified for 12 cycles with Phusion DNA polymerase (New England Biolabs). Each sample was sequenced on a single lane of an Illumina GAIIX sequencer (see File S1 for further details).

Comparative genomics

The restriction enzyme *Sbf*I cuts frequently in coding sequences; for example, the sequenced zebrafish genome has 26,948 *Sbf*I recognition sites (giving 53,896 RAD tags), with

6010 (11%) of those occurring in protein-coding genes; for stickleback, the figure is even higher, at 16%. To improve identification of coding sequences located near each RAD tag, we constructed paired-end contigs by randomly shearing *Sbf*I-digested DNAs to obtain fragments of different length, all of which had a RAD tag at one end, and subjected them to paired-end sequencing (see File S1 for further details). Sequence from the first end defined the RAD-tag marker, while sequence from the paired ends, which occurred at many different distances from each tag due to random shearing, provided a few hundred base pairs of contiguous sequence located a few hundred base pairs from each enzyme cutting site (Figure S2A) (see also Etter *et al.* 2011).

BLASTx searches of zebrafish, stickleback, and human genomes (Ensembl v56), using an *e*-value cutoff of 1*e*-5 allowed the association of gar paired-end sequences and RNA-seq contigs to highly similar annotated sequences in teleosts. The annotation of each coding sequence that aligned on the gar map was manually verified. We constructed Oxford grids (Edwards 1991) by lining up all 974 gar coding markers in their genomic order along each gar chromosome displayed in numerical order on the horizontal axis and then plotting the position of the human ortholog on the human karyotype displayed along the vertical axis.

To quantify syntenic divergence, we selected a group of 588 zebrafish genes, 573 stickleback genes, and 486 human genes with mapped gar orthologs. We counted the number of chromosomes in human or gar that contained orthologs of genes on each teleost chromosome and the number of chromosomes in human that contained orthologs of genes on each gar chromosome, normalized to the number of chromosomes in each species (zebrafish, 25; stickleback, 21; gar, 29; and human, 23) and evolutionary divergence times (Figure 1) (Inoue *et al.* 2005), and plotted results on a distance tree. To avoid genome duplication effects, we compared genomes unidirectionally from teleost to gar and human.

Results

Constructing the gar map

To test, on a genome-wide scale, whether spotted gar lineage diverged from the teleost lineage before the TGD (Hoegg *et al.* 2004; Crow *et al.* 2006) and to challenge the controversial hypothesis that chromosome rearrangements accelerated after the TGD (Comai 2005; Semon and Wolfe 2007; Hufton *et al.* 2008), we netted from nature a male and a female spotted gar and mated them by *in vitro* fertilization. We isolated genomic DNA from parents and from 94 of their 2-week old offspring to form our F_1 mapping panel. Markers heterozygous in the female provided a female meiotic map and markers heterozygous in the male parent provided a male map, while markers segregating in both parents were mapped as an F_2 family and allowed the construction of a combined map (Figure 2, A and B). Each parental library provided 5 million sequences to identify the universe of RAD tags present in the cross. In all, gar had ~33,000 *Sbf*I restriction cut sites, which provided ~65,000 total tags because tags extend in each direction from each cut site. Each of the 94 F_1 progeny provided ~1 million sequences to ascertain their genotypes (Table S1).

To automate genotype calling, we utilized Stacks software (Catchen *et al.* 2011). In brief, Stacks assembled RAD tags into stacks of identical sequence (Figure 2C), compared stacks pairwise (Figure 2D), and merged stacks into loci, defined as stacks of average sequencing depth that differ by fewer than three nt (Figure 2E). Stacks distinguished sequencing errors from polymorphisms using a maximum likelihood framework (Hohenlohe *et al.* 2010), compared loci of offspring and parents, called genotypes, provided a web interface for interrogating sequences and genotypes, and exported genotypes into JoinMap mapping software (Van Ooijen 2006). Of ~65,000 total tags, 15,076 were polymorphic, and of those, 8406 tags mapped to the male map, the female map, or both. Polymorphic markers that did not appear in the final maps failed due to sequence ambiguities that decreased the number of map cross progeny scored for those markers below criterion (85 individuals). Because the number of RAD-tag markers greatly exceeded the limits of JoinMap, we arbitrarily selected 4551 markers plus all protein-coding loci for the final map.

JoinMap assigned 5466 markers to 29 linkage groups, similar in number to the 28 chromosomes of the closely related longnose gar (*L. osseus*) (Rab *et al.* 1999). Total map length was 1988 cM. Loc1, the longest *L. oculatus* linkage group, had 598 markers in 84 cM; Loc29, the shortest linkage group, had 44 markers in 62 cM; and Loc15 had 145 markers in 52 cM with 34 markers in protein-coding regions (Figure 2G). Figure S1 shows the complete map. A total of 656 markers were polymorphic in both the male and the female and hence were shared between the male and the female maps; these markers showed the same order and location in both maps, thus verifying map validity.

Comparative genomics

Coding sequences are required for comparative genomics, and although some of the mapped RAD-tag sequences represented coding regions [121 of 8406 RAD tags (1.4%) when compared to human], we utilized two strategies to increase the number of mapped coding sequences (Figure S2).

First, we randomly sheared *Sbf*I-digested DNAs to obtain fragments of different length with a RAD tag at one end and subjected them to paired-end sequencing. Sequencing from the first end defined the RAD-tag marker and hence genomic location, while sequencing from the paired end, which occurred at many different distances from each tag due to random shearing, provided a few hundred base pairs of sequence located a few hundred base pairs from each enzyme cut site (Figure S2A). This procedure greatly increased the length of sequence associated with each RAD tag and increased the chance of finding sequence in coding regions.

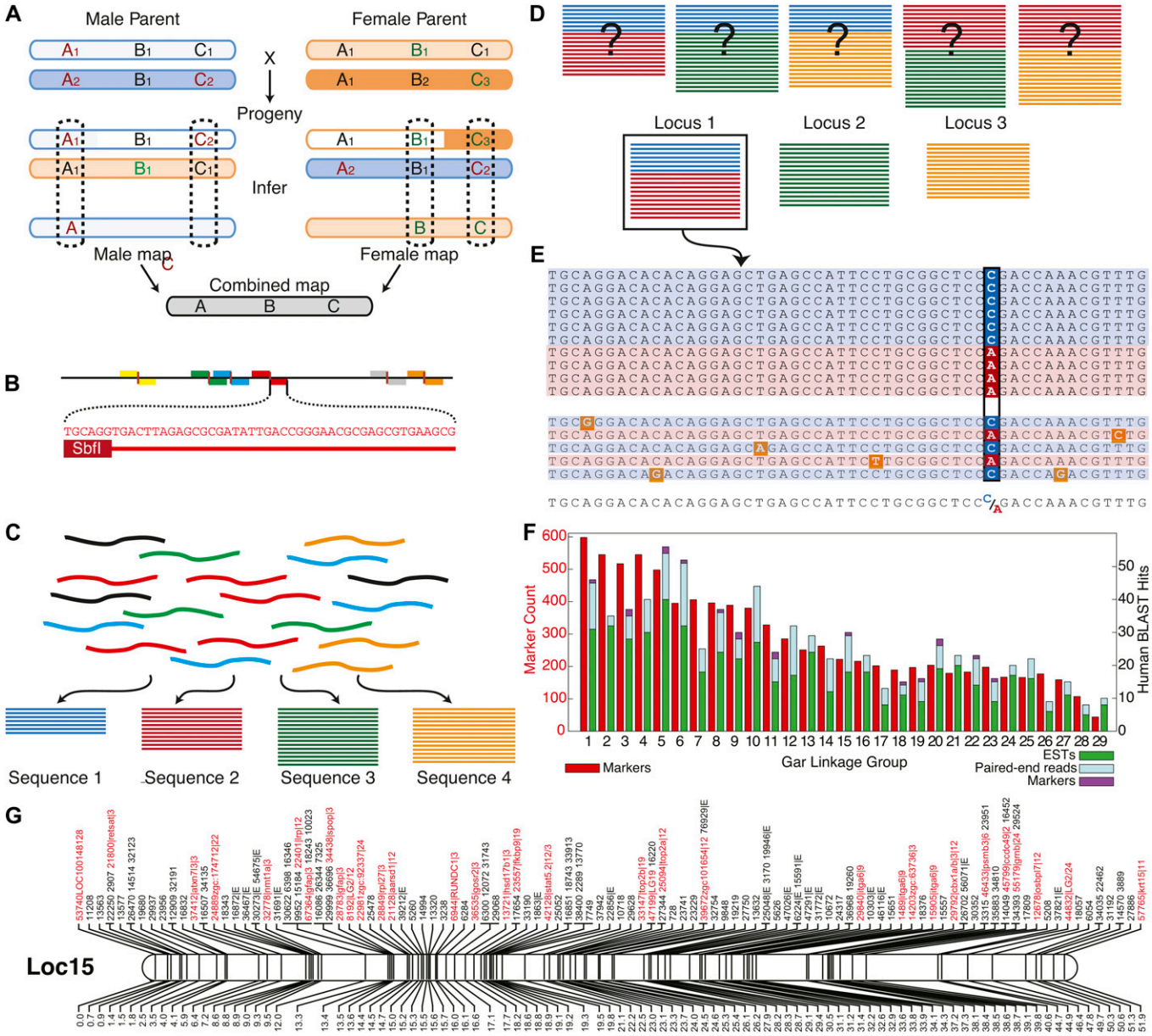

Figure 2 Generating a RAD-tag genetic map. (A) Markers heterozygous in one parent and homozygous in the other (*A* in the male and *B* in the female) map as a backcross. Markers heterozygous in both parents with one shared allele (*C*) associate sex-specific maps into a combined map. (B) Digesting DNA with a restriction enzyme and sequencing in both directions from the cut site yields RAD tags. (C) Stacks software sorts tags into stacks of identical sequences. (D) Comparing stacks pairwise joins stacks into loci that differ at fewer than three nt. (E) Errors are sporadic, but alleles are identified statistically when two alternative nucleotides at a single position are each present in a significant number of reads. (F) Marker number varied among the 29 linkage groups (red columns, scale on the left), as did the number of coding markers (right scale), including ESTs (green), paired-end reads (blue), and RAD tags in coding regions (purple). (G) Linkage group Loc15 from the combined map showing 145 mapped markers, 34 of which are coding (red).

The second step to increase mapped coding genes was to conduct a small gar transcriptomics project by RNA-seq (Pan *et al.* 2008). Our gar reference transcriptome comes from mRNA isolated from one mature embryo and from the head of another individual larva a few days older, both from the F_1 map cross population and thus they share polymorphisms with mapped coding RAD tags. For each sample, one lane of sequencing gave 62.5 million raw paired-end reads 60 nt long with an insert length of 350 bp. Quality filtering and trimming for quality left 38.5 million paired-end reads and 6.7 million single-end reads. Velvet (Zerbino and Birney 2008) assembled 18.3 million of these reads into the final transcriptome assembly using a k-mer coverage estimate of 38x (based on a comparison to the known zebrafish transcriptome) and with a coverage cutoff of four. Figure S3 plots the number of contigs *vs.* contig length for this transcriptome. BLASTn searches against the gar reference transcriptome, using as query mapped RAD tags and their several hundred base pairs of paired-end contigs, identified an additional 691 mapped markers with near perfect

identity to EST contigs. A BLASTn search of all 65,000 RAD tags, whether polymorphic or not, against our gar reference transcriptome identified 9086 (14%) that hit an EST, according to the criteria that the alignment spanned at least 70% of the query, hit the EST contig with an *e*-value of 1*e*-20 or less, and had a top BLAST hit with a raw score that was at least an order of magnitude better than the second best hit; these criteria accommodate exon/intron boundaries and allow for a few mismatches due to sequencing error or polymorphism. Of RAD tags that aligned to the transcriptome, 1327 (14.6%) contained polymorphisms, and of those, 945 were placed on the final map; the rest (382) were excluded because they were associated with RAD tags that were not scored in the minimum number of progeny that we had arbitrarily set as necessary to appear on the map.

BLASTx searches connected mapped markers, paired-end sequences, and ESTs to annotated genes in the sequenced genomes of human (*Homo sapiens*), stickleback (*Gasterosteus aculeatus*), and zebrafish (*Danio rerio*). We chose stickleback because its genome is particularly well assembled and zebrafish because it represents a lineage that diverged basal to other sequenced teleosts. Manual curation associated 891 and 900 mapped markers to zebrafish and stickleback orthologs (or co-orthologs) and 777 markers to human orthologs; in total, the analysis provided 974 gar markers with a putative ortholog in at least one of the three other genomes (Figure 2F).

The assignment of mapped markers to coding genes provided an additional test of map validity. A total of 58 coding genes contained more than one map marker each. Importantly, in each case all markers within the same gene mapped to the same bin or neighboring bins, as expected if the map were accurate.

With nearly 1000 coding markers distributed over the genome, we could perform comparative genomic analyses. We constructed Oxford grids (Edwards 1991) that lined up all 974 gar markers in their genomic order along the horizontal axis and plotted the position of each human ortholog on the human genome displayed along the vertical axis (Figure 3A). Results showed, first, that each segment of a human chromosome tends to be represented on just one gar chromosome, rather than on two chromosomes as in teleosts (Postlethwait *et al.* 1998; Woods *et al.* 2000; Jaillon *et al.* 2004). For example, the short arm of *H. sapiens* chromosome 5 (Hsa5p) is orthologous to the upper (right) part of Loc6, while the proximal part of the long arm of human chromosome 5 (Hsa5q) is orthologous to the proximal (left) part of Loc2, and the distal tip of Hsa5q is orthologous to the distal tip of Loc9 (Figure 3, A and B). Likewise, all orthologs of Hsa4 genes were found on Loc4 (Figure 3A). Although inversions in the gar or human lineages or both have rearranged regions orthologous to Hsa17, analysis shows that large regions are represented on a single gar chromosome (Figure 3C). Reciprocally, the upper (right) portion of Loc10 is orthologous only to Hsa17 and the lower part to Hsa1 or Hsa19 (Figure 3A). Likewise, the human orthologs of mapped Loc19 genes are located only on Hsa19q (Figure 3A). In contrast, each portion of a human chromosome tends to fall on two teleost chromosomes (Amores *et al.* 1998; Taylor *et al.* 2003; Jaillon *et al.* 2004; Kasahara *et al.* 2007).

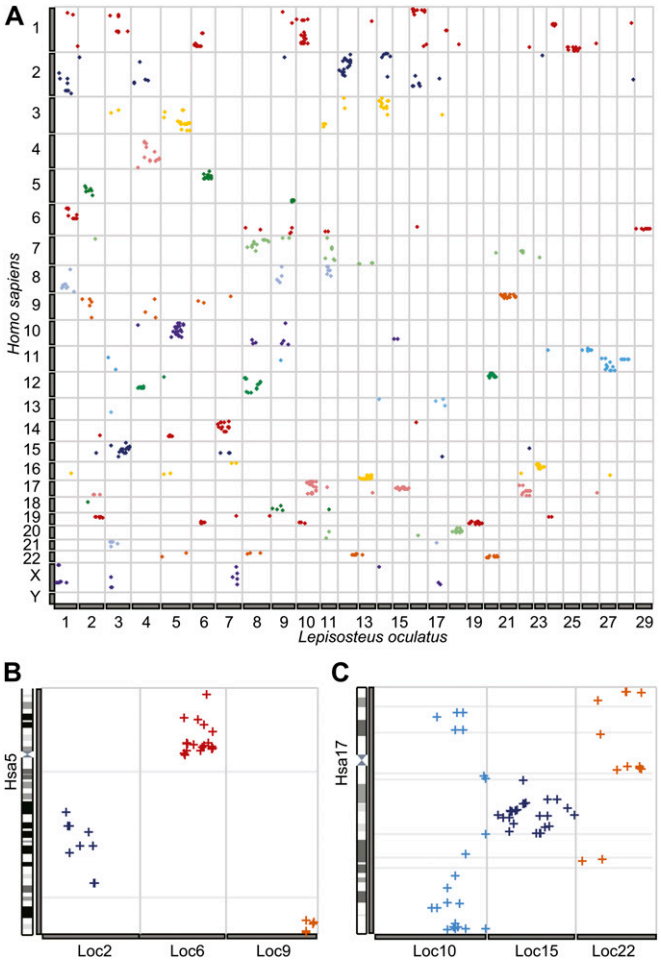

Figure 3 Comparative genomics of gar and human. (A) Oxford grid comparing human (*Homo sapiens*) and gar (*Lepisosteus oculatus*) genomes. Each symbol represents the position of an orthologous pair of genes in each genome. (B) Different portions of human chromosome 5 (Hsa5) correspond to nonoverlapping parts of different gar chromosomes (Loc2, Loc6, and Loc9). (C) Different portions of Hsa17 correspond to nonoverlapping parts of different gar chromosomes (Loc10, Loc15, and Loc22).

Second, as with human chromosomes, each portion of a gar chromosome is in general orthologous to parts of two different teleost chromosomes. For example, gar genes that mapped to the left part of Loc10 have orthologs distributed broadly over two stickleback chromosomes, LGVIII and LGIII, while genes located on the right portion of Loc10 have stickleback orthologs distributed along LGXI and LGV+LGIX (Figure 4A). In addition, the Loc10 set of Hsa17 orthologs occurs on two zebrafish linkage groups that were previously shown to be paralogous, Dre3 and Dre12 (Postlethwait *et al.* 1998) (Figure 4B). This includes the gar orthologs of human genes *GRIN2C* and *SDK2*, each of which has one co-ortholog on zebrafish Dre3 and the other co-ortholog on

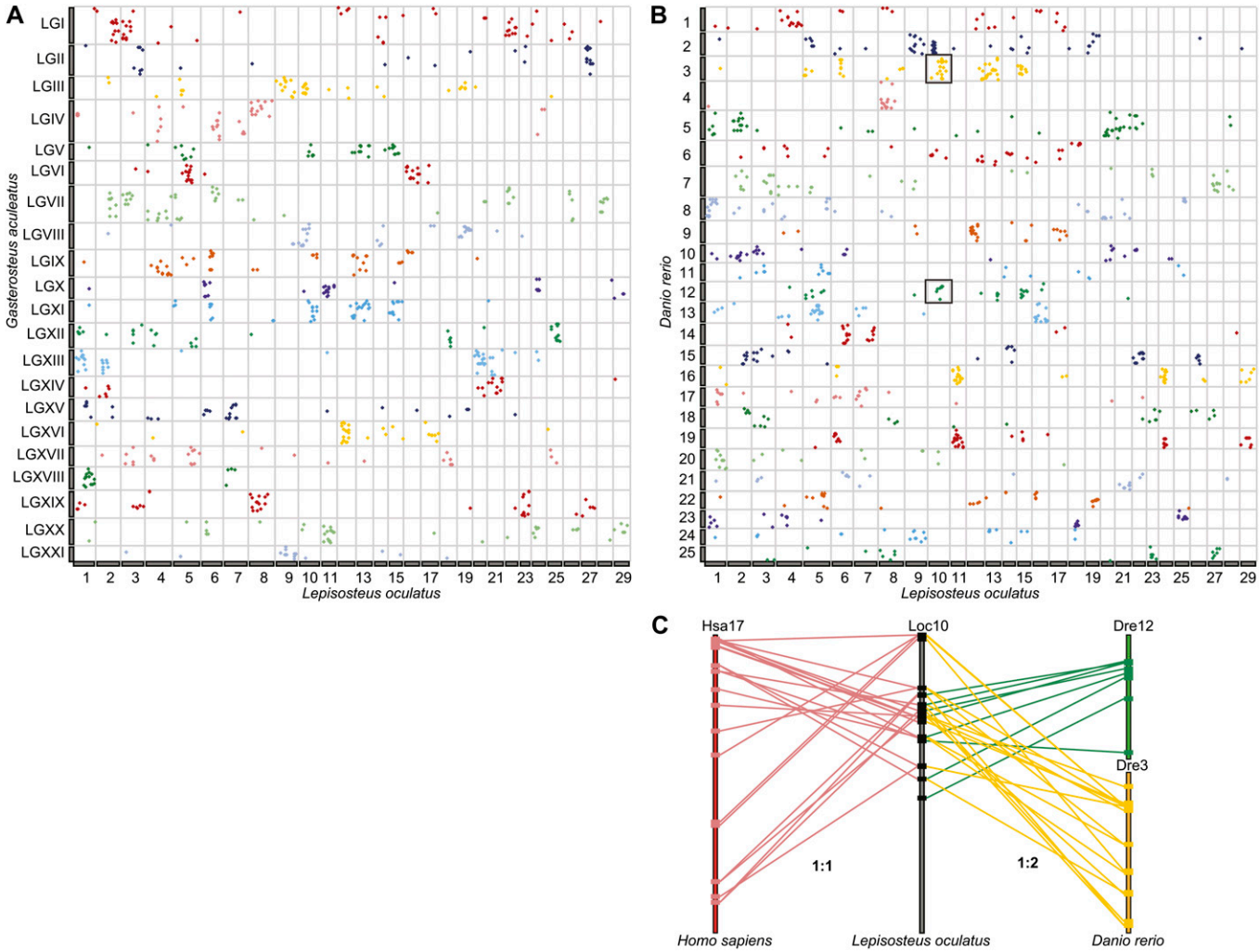

Figure 4 Comparative genomics of gar and teleosts. (A) Oxford grid comparing stickleback (*Gasterosteus aculeatus*) and gar (*Lepisosteus oculatus*) genomes. (B) Oxford grid comparing zebrafish (*Danio rerio*) and gar genomes. (C) Orthologs of Hsa17 genes were distributed on the upper portion of Loc10 but appeared on two different zebrafish chromosomes, Dre3 and Dre12, with duplicates of some genes occupying both zebrafish chromosomes.

Dre12 (Figure 4, B and C). Likewise, Loc19 is co-orthologous to zebrafish chromosomes Dre2 and Dre22 (Figure 4). This evidence shows on a genome-wide scale that gar and teleost lineages diverged before the TGD.

Teleost genome rearrangements

Data concerning the order of coding markers on the gar genetic map provide an opportunity to distinguish competing hypotheses for the explanation of the origin of teleost genome rearrangements. Under one hypothesis, chromosome rearrangements accelerated in the teleost lineage after the TGD (Comai 2005; Semon and Wolfe 2007); alternatively, most rearrangements had already occurred in the human or ray-fin lineage or both before the divergence of gar and teleost lineages and are thus mostly independent of genome duplication (Hufton *et al.* 2008). The first hypothesis predicts that gar and human genomes would be about equally rearranged with respect to teleost genomes. In contrast, the second hypothesis predicts that the architecture of teleost genomes would be more similar to the gar genome than to the human genome. We quantified syntenic divergence and normalized to evolutionary divergence times. Results showed that human and gar genomes clustered together separated from the two teleost genomes by a long branch (Figure 5). This finding shows that syntenic rearrangements accelerated after the divergence of gar and teleost lineages but before the divergence of stickleback and zebrafish lineages; this result is predicted by the hypothesis that the TGD facilitated the fixation of chromosome translocations in the teleost lineage.

Discussion

We present here a strategy that provides to nonmodel species a rapid and economical method for constructing dense, coding-rich meiotic maps from the offspring of individual wild-caught parents. We used this approach to rapidly produce the largest published meiotic map for any

fish, to our knowledge, containing >8000 total markers and nearly 1000 protein-coding markers. Importantly, this map was achieved for orders of magnitude less expense than other genome projects. The new map showed—on a genome-wide scale—that gar and teleost lineages diverged before the teleost genome duplication. Phylogenetic analyses suggest that sturgeons (*Acipenser sp*.), gar (*Lepisosteus sp*.), and bowfin (*Amia calva*) occupy a clade of ancient ray-fin fish that diverged from the teleost lineage after the divergence of the bichir (*Polypterus sp*.) lineage (Hoegg *et al.* 2004; Inoue *et al.* 2005; Crow *et al.* 2006; Katsu *et al.* 2008; Salaneck *et al.* 2008). Among species occupying this pre-TGD clade, spotted gar appears to be the most suitable for studies of development, genomics, and physiology. Spotted gar are locally plentiful in North America from Louisiana to Ontario, are amenable to *in vitro* fertilizations in the laboratory (in contrast to bowfin), do not have multiple rounds of genome duplication as do many species of sturgeons and paddlefish (Birstein and Desalle 1998), and have relatively small genomes compared to species such as bichir, a basally diverging ray-fin fish (Hardie and Hebert 2004). Furthermore, gar can be raised to adulthood in the laboratory, they produce thousands of eggs in a single spawn, their embryos are suitable for *in situ* hybridization analyses of gene expression, and their large size provides substantial material for biochemical and physiological analyses.

We conclude that spotted gar is the species of choice to serve as an experimentally accessible outgroup to teleosts to help infer ancestral, preduplicated functions of genes duplicated in teleosts. For example, if one member of a pair of teleost gene duplicates has a function that is not found in humans and other tetrapods, for example melanocyte adaptation to background coloration (Braasch and Postlethwait 2011; Zhang *et al.* 2010a), it is unclear whether that function was ancestral in bony fish and was lost in the human lineage or was lacking in the common ancestor of all bony fish and was newly evolved in the teleost lineage, perhaps as a neofunctionalization event (Force *et al.* 1999) after the TGD. Investigation of function in gar embryos could help resolve these types of questions.

Comparative genomic analyses conducted with the gar map revealed substantial conservation of synteny between human and gar that is consistent with the model that the TGD accelerated the loss of ancestral syntenies, and thus fails to rule out the hypothesis that whole-genome duplication plays a role in promoting syntenic rearrangements. Several theoretical concerns suggest possible mechanisms by which genome duplication could accelerate chromosomal rearrangements. The duplication of homologous coding elements would provide more substrates for illegitimate recombination between homeologous (paralogous) chromosomes and thereby stimulate the chromosome translocations that disturb conserved syntenies (Comai 2005). In addition, natural selection would favor rearranged karyotypes that reduce meiotic pairing between homeologous chromosomes because reduced pairing of homeologs would decrease the rate of aneuploid offspring, thereby improving fitness and thus tending to preserve these rearrangements during evolution (Comai 2005).

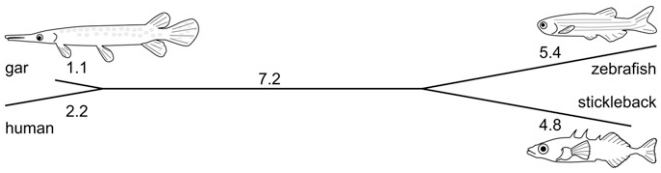

Figure 5 Syntenic comparisons of two teleosts (zebrafish and stickleback) to gar and human. Branch lengths are proportional to conserved syntenies estimated by counting the average number of gar (or human) chromosomes containing orthologs of genes on a teleost chromosome divided by the number of chromosomes in gar (or human) normalized to divergence times in hundreds of millions of years (Table S2). Branch lengths between gar and human are much shorter than lengths between either of these two species and teleost genomes.

The mapping strategy we develop here is potentially broadly applicable to nonmodel species, capitalizing on the power of massively parallel sequencing to reveal genetic polymorphisms and to accelerate transcriptome analysis. For this RAD-tag strategy to work well, a species should (1) have access to a clutch of 20 or preferably more individuals from a single female taken from nature to make a female map or from several females fertilized by the same male to make a male map (see Figure S4), (2) show sufficient heterozygosity to provide polymorphisms in RAD-tag sequences (note that increasing reads from 80 to 100 or 150 nt would increase the rate of capturing polymorphisms), (3) possess a genome that provides appropriate numbers of RAD tags (using a restriction enzyme that recognizes a 6-bp site would increase the number of tags, which would be useful for a small genome, and using an enzyme that recognizes a 10-bp sequence would decrease the number of tags, which would be useful for a large genome or for situations in which fewer markers were required to answer questions), and (4) provide sufficient material for RNA-seq to help identify coding sequences contained in RAD tags. While species like kiwis (*Apteryx australis*) and elephants that produce a single offspring each season would not be so favorable for this approach, it should work for thousands of other species of interest for physiology and evolution. RAD-tag meiotic mapping provides an inexpensive and rapid way for individual research laboratories to query genomes at nodes in the tree of life that interest them and to map naturally occurring genetic variants.

Although the first linkage maps appeared nearly a century ago (Sturtevant 1913), meiotic maps remain a critical tool for understanding the mechanisms of development, physiology, and evolution. In fact, in the age of whole-genome sequencing, dense, sequence-based genetic maps assume an even more important role to efficiently order the often thousands of unassembled genomic contigs resulting from genome sequencing projects (Lewin *et al.* 2009). For many species, meiotic maps using our strategy would be more rapid and far less costly than the construction of tiled, fingerprinted BACs and do not require the specialized skill sets

and materials needed for the development of radiation hybrid panels. The density of our meiotic map—an average of six polymorphic markers per megabase, or nearly one per BAC clone—would help assemble the thousands of contigs typically produced in today's genome sequencing projects. RAD-tag mapping, coupled with RNA-seq and low-coverage whole-genome sequencing by next-generation technologies, can liberate nonmodel organisms from the prison of genomic ignorance.

Acknowledgments

We thank W. Cresko, E. Johnson, I. Braasch, J. Smith, S. Bassham, and T. Titus for helpful conversations. This work was supported by National Institutes of Health (NIH)/National Center for Research Resources grant R01RR020833 to J.H.P, a Golden Ranch Plantation grant to A.F. and Q.F, an NIH National Research Service Award Ruth L. Kirschstein postdoctoral fellowship 1F32GM095213-01 to J.C., and National Science Foundation grant DEB-1025212 to W. Cresko.

Literature Cited

Albertson, R. C., W. Cresko, H. W. Detrich, 3rd, and J. H. Postlethwait, 2009 Evolutionary mutant models for human disease. Trends Genet. 25: 74–81.

Albertson, R. C., Y. L. Yan, T. A. Titus, E. Pisano, M. Vacchi *et al.*, 2010 Molecular pedomorphism underlies craniofacial skeletal evolution in Antarctic notothenioid fishes. BMC Evol. Biol. 10: 4.

Amores, A., A. Force, Y. L. Yan, L. Joly, C. Amemiya *et al.*, 1998 Zebrafish hox clusters and vertebrate genome evolution. Science 282: 1711–1714.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376.

Benton, M. J., and P. C. Donoghue, 2007 Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24: 26–53.

Birstein, V. J., and R. DeSalle, 1998 Molecular phylogeny of Acipenserinae. Mol. Phylogenet. Evol. 9: 141–155.

Blacklidge, K. H., and C. A. Bidwell, 1993 Three ploidy levels indicated by genome quantification in Acipenseriformes of North America. J. Hered. 84: 427–430.

Braasch, I., and J. H. Postlethwait, 2011 The teleost agouti-related protein 2 gene is an ohnolog gone missing from the tetrapod genome. Proc. Natl. Acad. Sci. USA 108: E47–E48.

Catchen, J. M., A. Amores, P. A. Hohenlohe, W. A. Cresko, and J. H. Postlethwait, 2011 *Stacks*: building and genotyping loci *de novo* from short-read sequences. G3: Genes, Genomes, Genetics 1: 171–182.

Chintamanani, S., S. H. Hulbert, G. S. Johal, and P. J. Balint-Kurti, 2010 Identification of a maize locus that modulates the hypersensitive defense response, using mutant-assisted gene identification and characterization. Genetics 184: 813–825.

Comai, L., 2005 The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6: 836–846.

Crow, K. D., P. F. Stadler, V. J. Lynch, C. Amemiya, and G. P. Wagner, 2006 The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol. Biol. Evol. 23: 121–136.

Eastman, J. T., and A. L. DeVries, 1981 Buoyancy adaptations in a swim-bladderless Antarctic fish. J. Morphol. 167: 91–102.

Edwards, J. H., 1991 The Oxford Grid. Ann. Hum. Genet. 55: 17–31.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson, 2011 Local de novo assembly of RAD paired-end contigs using short sequencing reads. PLoS One 6**:** e18561.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.

Grattapaglia, D., and R. Sederoff, 1994 Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudotestcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137.

Hardie, D., and D. Hebert, 2004 Genome-size evolution in fishes. Can. J. Fish. Aquat. Sci. 61: 1636–1646.

Hobert, O., 2010 The impact of whole genome sequencing on model system genetics: get ready for the ride. Genetics 184: 317–319.

Hoegg, S., H. Brinkmann, J. S. Taylor, and A. Meyer, 2004 Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J. Mol. Evol. 59: 190–203.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.

Hufton, A. L., D. Groth, M. Vingron, H. Lehrach, A. J. Poustka *et al.*, 2008 Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res. 18: 1582–1591.

Inoue, J. G., M. Miya, K. Tsukamoto, and M. Nishida, 2003 Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol. Phylogenet. Evol. 26: 110–120.

Inoue, J. G., M. Miya, B. Venkatesh, and M. Nishida, 2005 The mitochondrial genome of Indonesian coelacanth Latimeria menadoensis (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene 349: 227–235.

Jaillon, O., J. M. Aury, F. Brunet, J. L. Petit, N. Stange-Thomann *et al.*, 2004 Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431: 946–957.

Jeffery, W. R., 2009 Chapter 8. Evolution and development in the cavefish Astyanax. Curr. Top. Dev. Biol. 86: 191–221.

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.

Katsu, Y., S. Kohno, S. Hyodo, S. Ijiri, S. Adachi *et al.*, 2008 Molecular cloning, characterization, and evolutionary analysis of estrogen receptors from phylogenetically ancient fish. Endocrinology 149: 6300–6310.

Kucuktas, H., S. Wang, P. Li, C. He, P. Xu *et al.*, 2009 Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics 181: 1649–1660.

Lampert, K. P., C. Schmidt, P. Fischer, J. N. Volff, C. Hoffmann *et al.*, 2010 Determination of onset of sexual maturation and mating behavior by melanocortin receptor 4 polymorphisms. Curr. Biol. 20: 1729–1734.

Lewin, H. A., D. M. Larkin, J. Pontius, and S. J. O'Brien, 2009 Every genome sequence needs a good map. Genome Res. 19: 1925–1928.

Li, W., S. Royer, and R. T. Chetelat, 2010 Fine mapping of ui6.1, a gametophytic factor controlling pollen-side unilateral incompatibility in interspecific solanum hybrids. Genetics 185: 1069–1080.

Long, W. L., and W. W. Ballard, 2001 Normal embryonic stages of the longnose gar, Lepisosteus osseus. BMC Dev. Biol. 1: 6.

Meierjohann, S., and M. Schartl, 2006 From Mendelian to molecular genetics: the Xiphophorus melanoma model. Trends Genet. 22: 654–661.

Miller, M. R., T. S. Atwood, B. F. Eames, J. K. Eberhart, Y. L. Yan *et al.*, 2007a RAD marker microarrays enable rapid mapping of zebrafish mutations. Genome Biol. 8: R105.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson, 2007b Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17: 240–248.

Moore, J. L., L. M. Rush, C. Breneman, M. A. Mohideen, and K. C. Cheng, 2006 Zebrafish genomic instability mutants and cancer susceptibility. Genetics 174: 585–600.

Nakatani, Y., H. Takeda, Y. Kohara, and S. Morishita, 2007 Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17: 1254–1265.

Near, T. J., S. K. Parker, and H. W. Detrich, 3rd, 2006 A genomic fossil reveals key steps in hemoglobin loss by the antarctic icefishes. Mol. Biol. Evol. 23: 2008–2016.

Okada, M., C. Lanzatella, M. C. Saha, J. Bouton, R. Wu *et al.*, 2010 Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. Genetics 185: 745–760.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, 2008 Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40: 1413–1415.

Postlethwait, J., A. Amores, W. Cresko, A. Singer, and Y. L. Yan, 2004 Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 20: 481–490.

Postlethwait, J. H., Y.-L. Yan, M. Gates, S. Horne, A. Amores *et al.*, 1998 Vertebrate genome evolution and the zebrafish gene map. Nat. Genet. 18: 345–349.

Rab, P., M. Rabova, K. M. Reed, and R. B. Phillips, 1999 Chromosomal characteristics of ribosomal DNA in the primitive semionotiform fish, longnose gar Lepisosteus osseus. Chromosome Res. 7: 475–480.

Salaneck, E., T. A. Larsson, E. T. Larson, and D. Larhammar, 2008 Birth and death of neuropeptide Y receptor genes in relation to the teleost fish tetraploidization. Gene 409: 61–71.

Sanetra, M., F. Henning, S. Fukamachi, and A. Meyer, 2009 A microsatellite-based genetic linkage map of the cichlid fish, Astatotilapia burtoni (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. Genetics 182: 387–397.

Schartl, M., B. Wilde, J. A. Laisney, Y. Taniguchi, S. Takeda *et al.*, 2010 A mutated EGFR is sufficient to induce malignant melanoma with genetic background-dependent histopathologies. J. Invest. Dermatol. 130: 249–258.

Semon, M., and K. H. Wolfe, 2007 Rearrangement rate following the whole-genome duplication in teleosts. Mol. Biol. Evol. 24: 860–867.

Smith, O. A., 2006 Reproductive potential and life history of spotted gar *Lepisosteus oculatus* in the upper Barataria estuary, Louisiana, pp. 116 in *Marine and Environmental Biology*. Nicolls State University, Thibodaux, LA.

Streelman, J. T., J. F. Webb, R. C. Albertson, and T. D. Kocher, 2003 The cusp of evolution and development: a model of cichlid tooth shape diversity. Evol. Dev. 5: 600–608.

Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43–59.

Taylor, J., I. Braasch, T. Frickey, A. Meyer, and Y. Van De Peer, 2003 Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res. 13: 382–390.

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.

Valenzano, D. R., J. Kirschner, R. A. Kamber, E. Zhang, D. Weber *et al.*, 2009 Mapping loci associated with tail color and sex determination in the short-lived fish Nothobranchius furzeri. Genetics 183: 1385–1395.

Van Ooijen, J. W., 2006 JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, The Netherlands.

Woods, I. G., P. D. Kelly, F. Chu, P. Ngo-Hazelett, Y. L. Yan *et al.*, 2000 A comparative map of the zebrafish genome. Genome Res. 10: 1903–1914.

Zerbino, D. R., and E. Birney, 2008 Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829.

Zhang, C., Y. Song, D. A. Thompson, M. A. Madonna, G. L. Millhauser *et al.*, 2010a Pineal-specific agouti protein regulates teleost background adaptation. Proc. Natl. Acad. Sci. USA 107: 20164–20171.

Zhang, G., S. Hoersch, A. Amsterdam, C. A. Whittaker, J. A. Lees *et al.*, 2010b Highly aneuploid zebrafish malignant peripheral nerve sheath tumors have genetic alterations similar to human cancers. Proc. Natl. Acad. Sci. USA 107: 16940–16945.

Zuryn, S., S. Le Gras, K. Jamet, and S. Jarriault, 2010 A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics 186: 427–430.

Communicating editor: L. M. McIntyre

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.111.127324/-/DC1

Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication

Angel Amores, Julian Catchen, Allyse Ferrara, Quenton Fontenot, and John H. Postlethwait

DOI: 10.1534/genetics.111.127324

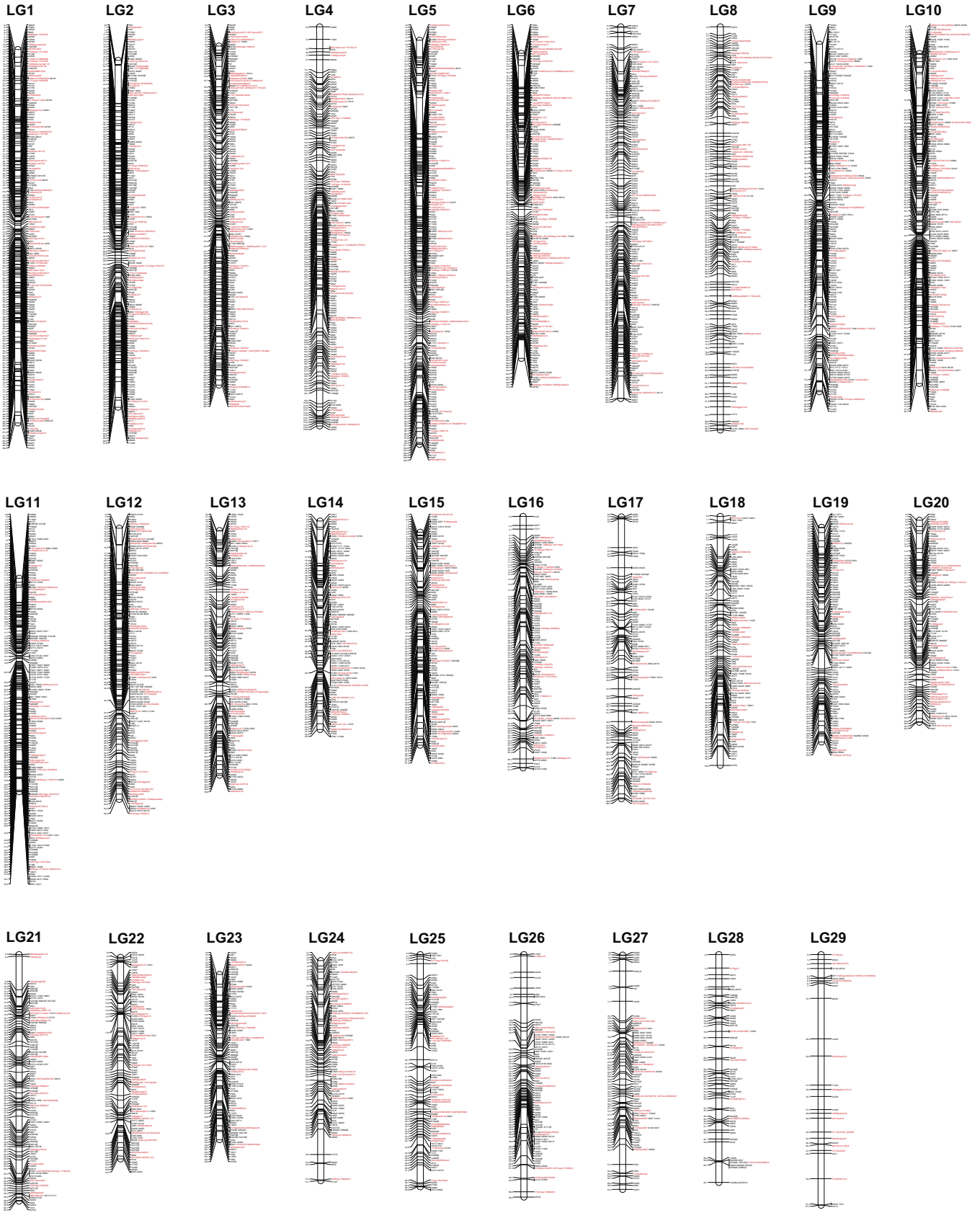

Figure S1 The spotted gar genetic map. Coding markers shown in red. Abbreviations: LG, linkage group.

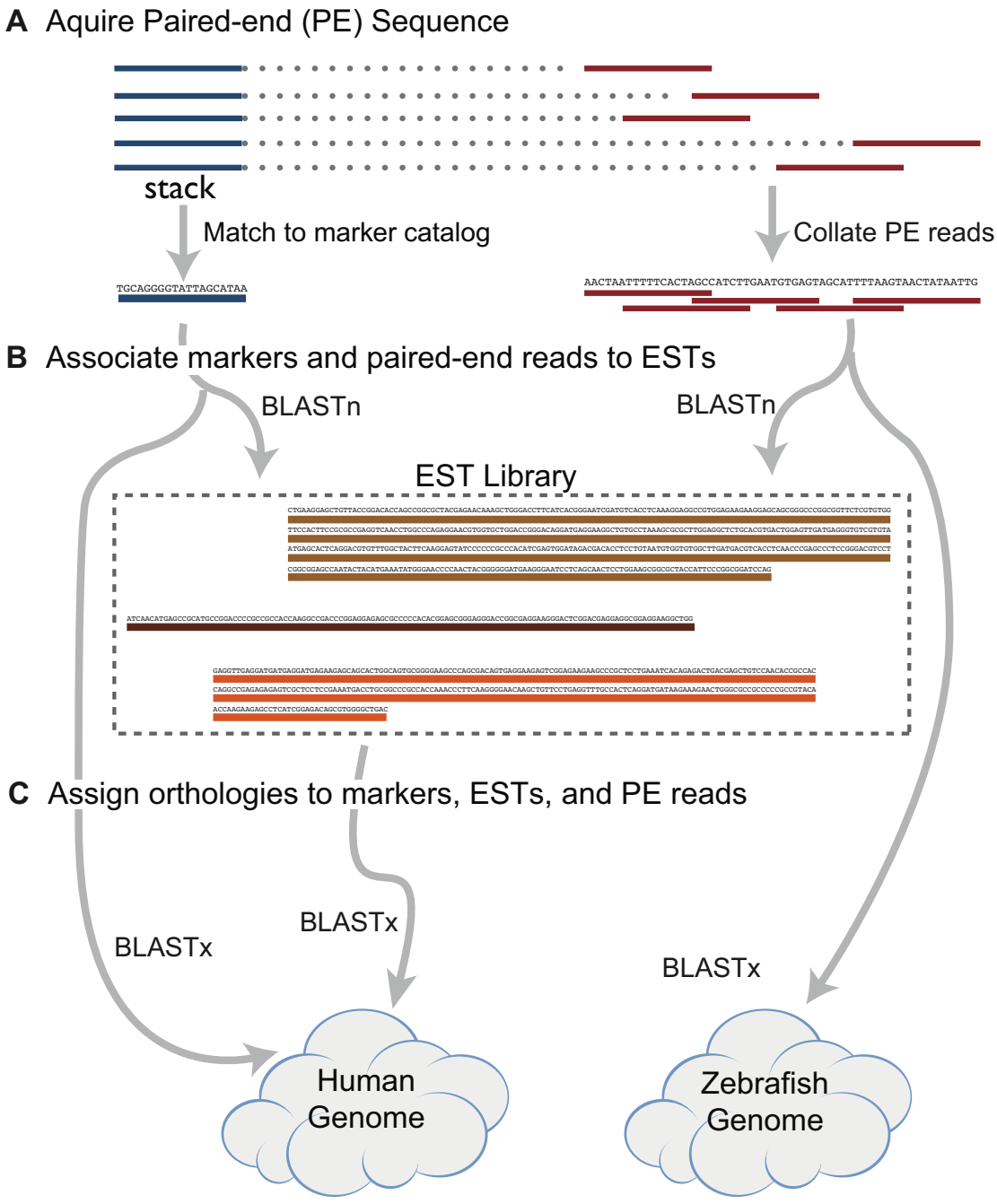

Figure S2 Strategy to associate coding sequences to mapped markers. A. Paired-end contig strategy. B. RNA-seq strategy. C. Assigning markers to coding sequences.

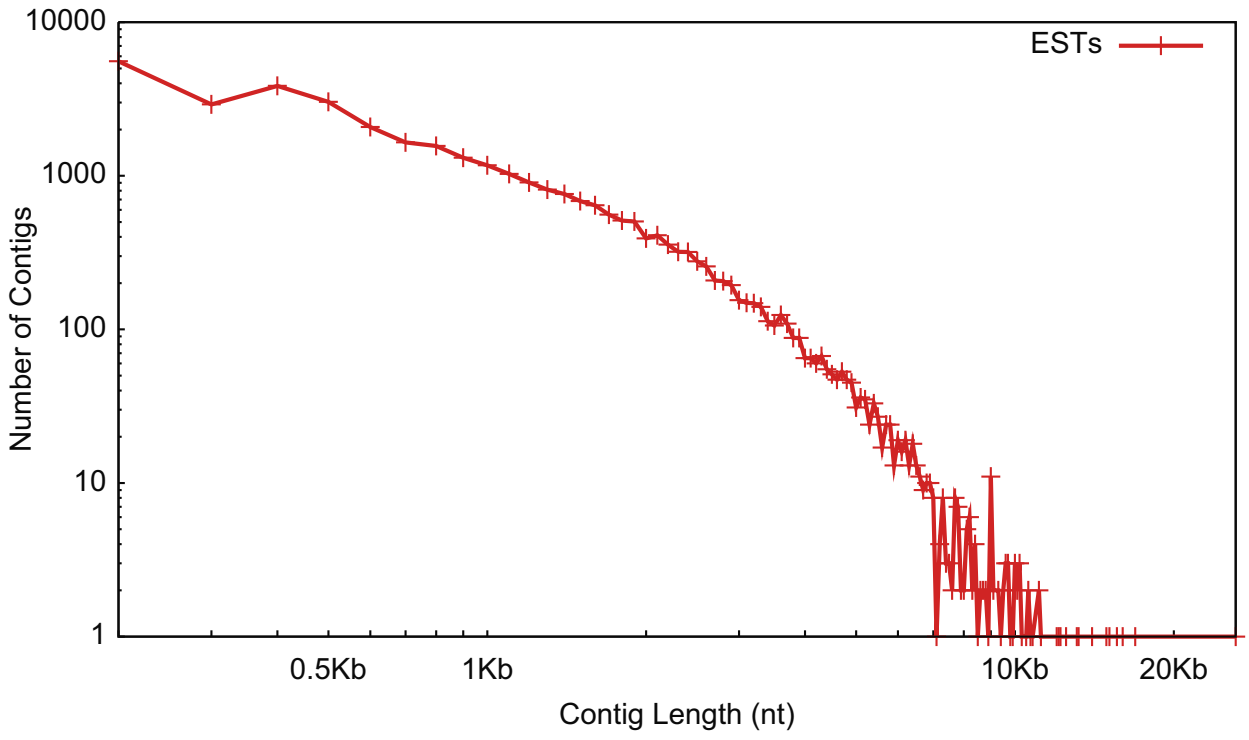

Figure S3 The RNA-seq assembly. Log of the number of contigs is plotted against the log of contig length in nucleotides.

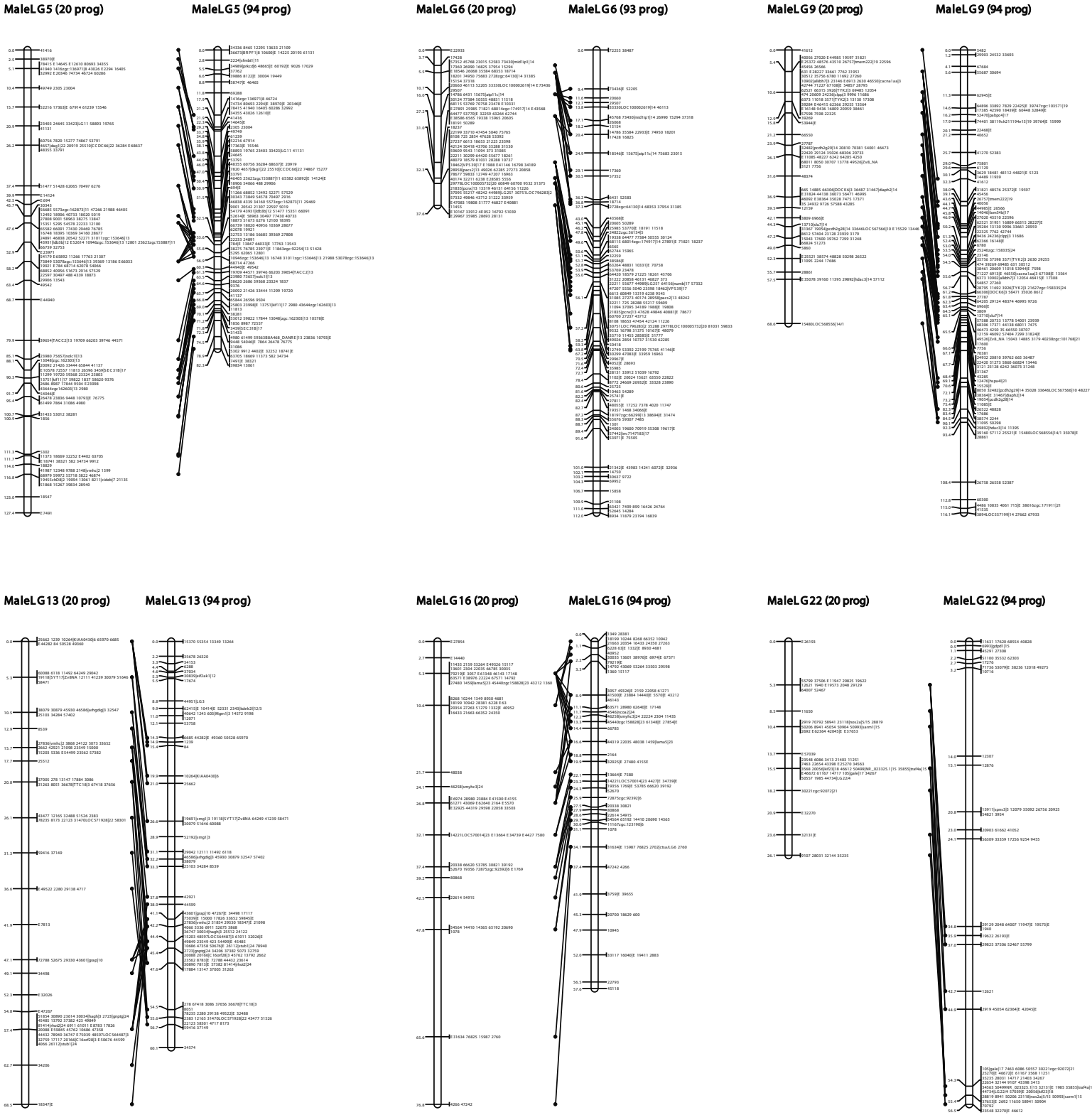

Figure S4 Useful maps can be created with as few as 20 map cross individuals. We randomly selected 20 individuals from our panel and constructed a map just from markers segregating in just those individuals. Inspection of six representative linkage groups constructed from that small number of progeny show that linkage groups based on 20 individuals were generally missing distal parts of chromosomes compared to those made with 94 individuals and occasionally, some markers were inverted. Nonetheless, major portions of linkage groups could be reconstructed with as few as 20 map cross individuals.

File S1

Methods for RAD-tags, paired-end contigs, and cDNA for RNA-seq

A. RAD-tags

1. Materials

1.1. DNA extraction and RNase A treatment

1. DNeasy Blood & Tissue Kit (Qiagen) (or similar)
2. RNaseA (Qiagen).

1.2. Restriction endonuclease digestion

1. Restriction enzyme (*Sbf*1-HF , New England Biolabs) : 20 U/µl
2. Clean, intact high-quality genomic DNA: >20 ng/µl.

1.3. P1 Adapter ligation

1. New England Biolabs (NEB) Buffer 2.
2. ATP (Epicentre): 100 m*M*.
3. P1 Adapter: 100 n*M.* A modified Solexa© adapter (2006 Illumina, Inc., all rights reserved). Prepare 100nM stocks of P1 in 1X NEB buffer #2 (or a buffer with a 50 mM final concentration of NaCl). For adapter annealing, prepare a 10 µ*M* stock of P1 adaptors, denature for 1 minute at 98C in PCR machine and decrease the temperature 1C per minute until 25C. Dilute to 100nM in 1X NEB #2 buffer and store in the freezer. Below, example barcoded *Sbf1* P1 adapter sequences. Asterisk denotes a phosphorothioate bond introduced to confer nuclease resistance to the double-stranded oligo, /5Phos/ denotes a phosphate group and "x" refers to barcode nucleotides.

Sbf1-P1 top oligo:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxTGC*A-3'

Sbf1-P1 bottom oligo:

5'-/5Phos/xxxxxAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'

4. T4 DNA Ligase (Epicentre): 10 U/µl.

1.4. Purification steps

1. DNA Clean & Concentrator-5 kit (ZYMO Research).

1.5. DNA shearing

1. Bioruptor, nebulizer or sonicator.

1.6. Size selection/agarose gel extraction

1. Agarose (Sigma or other manufacturer)
2. 0.5X TBE buffer
3. Loading Dye Solution (Fermentas or other manufacturer).
4. GeneRuler 100 bp DNA Ladder Plus (Fermentas or other manufacturer).
5. Razor blades.
6. MinElute Gel Purification Kit (Qiagen).

1.7. Perform end repair

1. End-It DNA End-Repair Kit (Epicentre).

1.8. 3´-dA overhang addition

1. NEB Buffer 2.
2. dATP (Fermentas): 10 m*M*
3. Exo-Minus Klenow DNA polymerase (Epicentre): 10 U/µl.

1.9. P2 Adapter ligation

1. New England Biolabs Buffer 2.
2. ATP: 100 m*M.*
3. P2 Adapter: 10 µ*M.* A modified Solexa© adapter (2006 Illumina, Inc., all rights reserved). Prepare 10 µ*M* stocks of PE (Paired End) adapter P2 in 1X NEB buffer #2 (or buffer with a 50 mM final concentration of NaCl). For adapter annealing, denature for 1 minute at 98C in PCR machine and decrease the temperature 1C per minute until 25C. Store in the freezer. Asterisk denotes a phosphorothioate bond and /5Phos/ denotes a phosphate group.

PE-P2- top oligo

5'-/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCAGAACAA-3'

PE-P2- bottom oligo

5'- CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T -3'

4. T4 DNA Ligase (Epicentre) 10 U/µl.

1.10. RAD tag Amplification/Enrichment

1. Phusion High-Fidelity DNA polymerase with HF Buffer (New England Biolabs).
2. dNTP 100 mM (Fermentas).
3. Modified Solexa© Amplification primer mix (2006 Illumina, Inc., all rights reserved): 10 µ*M.*

P5-forward primer: 5´- AATGATACGGCGACCACCGA -3´

P7-reverse primer: 3´- CAAGCAGAAGACGGCATACGA -3´

2. Methods

The protocol described below describes methods to prepare RAD-tag libraries for high-throughput Illumina sequencing (see also (BAIRD *et al.* 2008; HOHENLOHE *et al.* 2010; MILLER *et al.* 2007a; MILLER *et al.* 2007b). The P1 adapter contains forward amplification and Illumina sequencing primer sites, as well as a barcode for sample identification. The second adapter (PE-P2), can be used for Paired-End sequencing. For genetic mapping, each individual sample has a unique barcode.

2.1. DNA extraction and RNase A treatment

1. Extract genomic DNA using the DNeasy Blood & Tissue Kit (Qiagen) or a similar product that produces very pure, high molecular weight, RNA-free DNA. High-quality DNA is extremely important. Quantify the DNA using a Qubit (Invitrogen) fluorometer to get the most accurate concentration readings.

2.2. Restriction endonuclease digestion

1. Digest 500ng-1µg genomic DNA for each individual sample with *Sbf*1-HF restriction enzyme in a 50 µl reaction volume, following the manufacturers instructions. In a microcentrifuge tube combine:

   5.0 µl 10X NEB Buffer 4
   0.5 µl *Sbf*1-HF (20 U/µl)
   500 ng-1µg genomic DNA
   H_2O to 50.0 µl.
   Incubate at 37C for 1h

2. Heat-inactivate the restriction enzyme for 30 minutes at 65C. Allow reaction to cool slowly to ambient temperature (30-60 min). If the enzyme cannot be heat-inactivated, purify with a ZYMO column following manufacturer's instructions prior to ligation.

2.3. P1 Adapter ligation

1. This step in the protocol ligates barcoded, Sbf1 restriction-site overhang P1 adapters onto Sbf1 compatible ends of the genomic DNA digested in the previous step
2. To each inactivated digest, add:

   6.0 µl 10X NEB Buffer 2
   3.0 µl Barcoded P1 Adapter (100 n*M*)
   0.6 µl ATP (100 m*M*)
   0.5 µl T4 DNA Ligase (10 U/µl)

   60.0 µl total volume.

   Add P1 adapters to the reaction before the ligase to avoid re-ligation of the genomic DNA. Incubate reaction at room temperature for 60 min or overnight in the refrigerator.

5. Heat-inactivate T4 DNA Ligase for 30 min at 65° C. Allow reaction to cool slowly to ambient temperature (30 min).

2.4. Sample multiplexing

2. Combine barcoded samples at desired ratio (about 16-24 samples per pool). Use a 100-300 µl aliquot containing 1-3 µg DNA total and freeze the rest at -20° C.

For genetic linkage maps, the best way to determine how many samples can be combined is to perform a pilot experiment. The size and polymorphism level of the genome under investigation will determine how deeply the samples need to be sequenced. Multiplex parents with several progeny and sequence them as a single sample. Use *Stacks* (CATCHEN *et al.* 2011) (http://creskolab.uoregon.edu/stacks/) to determine the total number of tags per individual plus the total number of polymorphic markers at the read length sequenced and use these data to determine the optimal read length and optimal number of progeny to pool. In general, aim for at least 25x coverage in the post-processed *SbfI* loci. For example, 50,000 loci would require 1 million retained reads per individual. If the organism being investigated has a low level of polymorphism, longer reads will provide more SNPs and hence more mappable markers. Higher coverage is better and results in fewer wrong genotypes, which occur mostly in under sequenced heterozygotes, but 35x coverage is generally sufficient. Run a single lane of progeny at the selected pooling complexity to determine if the number of reads is close to the expected number. A low read count in a lane may result in many missing genotypes in addition to many wrong genotypes.

2.5. DNA shearing

1. Shear DNA samples to an average size of 200-500 bp. The 300-600bp fraction seems to work just as well, so use a bigger or smaller size fraction if needed.
2. Dilute ligation reaction to 100 µl in water and shear in Bioruptor 10 times for 30 sec on high following manufacturer's instructions (may need some optimization in different organisms). For paired-end sequencing and building mini-contigs from the paired ends, make three replicas from the sample and shear one for 4 cycles, another for 7 cycles and another for 10 cycles and pool them together after shearing.
3. Clean up sheared DNA sample(s) using a Zymo-5 column following manufacturer's instructions. Elute in 20 µl EB (Elution Buffer).

2.6. Size selection/agarose gel extraction

1. Run the entire sheared sample on a 1.25% agarose, 0.5X TBE gel for 45-60 min at 100 V, next to a 100 bp DNA Ladder for size reference.
2. Cut a slice of the gel spanning 200-500 bp (for paired ends separately cut the 200-400 fraction, 400-600 and 600-900 fractions and proceed with each separately). Extract DNA using MinElute Gel Purification Kit (QIAGEN) by melting the agarose gel slices in the supplied buffer at room temperature (18-22° C) with agitation for 30 min. Elute twice with 18 µl EB (Elution Buffer) each time into the same eppendorf tube.

2.7. Perform end repair

To the eluate from the previous step, add:

5 µl 10x buffer

5 µl dNTP mix (1m*M*)

5 µl ATP

1.0 µl End-It Enzyme Mix.

Incubate at room temperature (RT) for 45 min.

3. Purify with ZYMO-5 column. Elute twice with 22 µl EB (Elution Buffer) into eppendorf tube.

2.8. 3´-dA overhang addition

To the eluate from the previous step, add:

5.0 µl 10X NEB Buffer 2

1.0 µl dATP (10m*M*)

1.5 µl Klenow (exo$^-$) (10U/µl, Epicentre).

Incubate at 37ºC for 30 min. Allow reaction to cool slowly to ambient temperature (15 min).

3. Purify with ZYMO-5 column. Elute twice with 22 µl EB (Elution Buffer) into eppendorf tube.

2.9. PE-P2 Adapter ligation

1. This step in the protocol ligates the PE-P2 adapter, a "Y" adapter with divergent ends that contains a 3´ dT overhang, onto the ends of blunt DNA fragments with 3´ dA overhangs from the previous step.
2. To the eluate from previous step, add:

5.0 µl 10X NEB Buffer 2

1.0 µl PE-P2 Adapter (10 µ*M*)

1.0 µl ATP (100 m*M*)

0.5 µl T4 DNA Ligase (10U/µl, Epicentre)

Incubate reaction at room temperature for 60 min or overnight in the refrigerator

3. Purify with ZYMO-5 column. Elute twice with 26 µl EB (into same tube).

2.10. RAD tag Amplification/Enrichment

1. This step will perform high-fidelity PCR amplification on P1 and P2 adapter-ligated DNA fragments.
2. Quantify the DNA using a fluorometer (Qubit, Invitrogen) to get the most accurate concentration readings.

Perform PCR amplification to determine library quality. In thin-walled PCR tube, combine:

10 µl 5x HF buffer (NEB)

1 µl dNTPs (10 mM)

40-70 ng RAD library template (eluate from last step)

2.0 µl Solexa primer mix (P5+P7 primers, 10 µ*M*)

0.5 µl Phusion DNA polymerase (2 U/µl , NEB)

H_2O to 50 µl

Perform 12 cycles of amplification in thermal cycler:

30 sec 98° C

12x [10 sec 98° C, 30 sec 65° C, 30 sec 72° C],

5 min 72° C

hold 10° C.

Purify PCR reaction reaction with a ZYMO-5 column. Elute in 20 µl EB.

4. Load entire sample in 1X Orange Loading Dye on a 1.25% agarose, 0.5X TBE gel and run for 45 min at 100 V, next to 100 bp DNA size standard for size reference. Use a fresh razor blade to cut a slice of the gel spanning 250-550 bp (the amplified library will migrate at a slightly higher size range than the template). If using a smaller or larger size fraction, cut the amplified product corresponding to that size range. For paired ends cut the corresponding fraction according to the expected size range. Extract DNA using MinElute Gel Purification Kit following manufacturer's instructions. Melt agarose gel slices at room temperature in the supplied buffer. Elute in 20 µl EB.

5. Quantify the DNA using a Qubit (Invitrogen) fluorometer to get the most accurate concentration readings. Concentrations will range from 1-20 ng/µl. Determine the molar concentration of the library by examining the gel image and estimating the median size of the library smear. Use this number to calculate the molar concentration of the library.
6. For paired end sequencing, combine the libraries from the 200-400, 400-600 and 600-900 fractions at equal ratios.
7. Sequence libraries on Illumina Genome Analyzer following manufacturer's instructions.

B. Processing Illumina data *in silico* to recover *RAD-tag* loci and paired-end mini-contigs

The *Stacks* software package (CATCHEN *et al.* 2011) (http://creskolab.uoregon.edu/stacks/) can help to recover loci from map cross parents and progeny and to build paired-end mini-contigs associated with those loci. Briefly, *Stacks* identifies a locus in an individual by aligning sequenced Illumina reads adjacent to the restriction enzyme cut site. In the case of paired end reads, *Stacks* aligns the first end to form a locus and collates the paired end to create groups of reads associated with each locus. The collated sequences for each locus are fed into an assembler such as Velvet (ZERBINO and BIRNEY 2008) and the resulting assembled contigs are loaded into the Stacks database. *Stacks* can export the assembled paired-end mini-contigs associated with their upstream locus, and these sequences can be used to link loci to EST or genomic sequences from the same species or to identify orthologs in other species and perform conserved synteny analyses.

C. RNAseq Methods

1. Materials

1.1. Total RNA and mRNA isolation

1. RiboPure Kit (Ambion).
2. MicroPoly(A) Purist (Ambion).

1.2. 1st strand synthesis

1. Random Primers (hexamers) (3 µg/µl, Invitrogen).
2. High-quality mRNA: >10 ng/µl. (100 ng -1µg)
3. dNTP (Fermentas)
4. 5x First Strand Synthesis buffer (Invitrogen).
5. RNase inhibitor (Invitrogen or New England Biolabs).
6. Superscript III reverse transcriptase (Invitrogen).
7. RNase H (5 U/µl, New England Biolabs).

1.2. Second strand synthesis

1. Random Primers (hexamers) (3 µg/µl, Invitrogen).
2. New England Biolabs buffer #2 (as alternative use Klenow exo- buffer (Epicentre)
3. dNTP (Fermentas
4. Klenow exo- (10 U/µl, Epicentre) (alternatively use DNA polymerase I)

1.3. PE Adapter ligation

1. NEB Buffer 2.
2. ATP (Epicentre): 100 m*M*.
3. T-overhang PE (Paired End) Adapter: 10 µ*M.* A modified Solexa© adapter (2006 Illumina, Inc., all rights reserved). Prepare 10 µ*M* stocks of PE adapters in 1X NEB buffer #2 (or buffer with a 50 mM final concentration of NaCl). For adapter annealing, denature for 1 minute at 98C in PCR machine and decrease the temperature 1C per minute until 25C. Store in the freezer.

Below, example PE adapter sequences with no barcode (PEnoBC). Asterisk denotes a phosphorothioate bond introduced to confer nuclease resistance to the double-stranded oligo, "/5Phos/" denotes a phosphate group.

PEnoBC-top:

5´- ACACTCTTTCCCTACACGACGCTCTTCCGATC*T -3´

PEnoBC- bottom:

5´- /5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG -3´

4. T4 DNA Ligase (Epicentre): 10 U/µl.

1.4. Purification steps

1. DNA Clean & Concentrator-5 kit (ZYMO Research).

1.5. DNA shearing

1. Bioruptor, nebulizer or Branson sonicator 450.

1.6. Size selection/agarose gel extraction

1. Agarose (Sigma or other manufacturer)
2. 1X TBE
3. 6X Orange Loading Dye Solution (Fermentas).
4. GeneRuler 100 bp DNA Ladder Plus (Fermentas).
5. Razor blades.
6. MinElute Gel Purification Kit (Qiagen).

1.7. End repair

1. End-It DNA End-Repair Kit (Epicentre).

1.8. 3´-dA overhang addition

1. NEB Buffer 2.
2. dATP (Fermentas): 10 m*M*
3. Klenow Fragment (3´ to 5´ exo$^-$, Epicentre): 10 U/µl.

1.9. Library Amplification

1. Phusion High-Fidelity DNA polymerase with HF Buffer (NEB).
2. Modified Solexa© Amplification primer mix (2006 Illumina, Inc., all rights reserved): 10 µ*M.*

PEprimer1

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

PEprimer2

5'- CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'

2. RNA seq Methods

First strand synthesis

Combine:

1 µl Random Primers (3 µg/µl, Invitrogen)

1 µl 10 mM dNTPs (Fermentas)

11 µl purified mRNA or up to 1 µg (Ambion MicroPolyA-Purist)

Heat to 65˚C for 5 min., then ice

Collect contents at bottom of tube by brief centrifugation.

Add:

4 µl 5x First Strand Synthesis buffer (Invitrogen)

1 µl 0.1 mM DTT (Invitrogen)

1 µl RNase inhibitor (Invitrogen or New England Biolabs)

1 µl Superscript III reverse transcriptase (Invitrogen)

Mix by gentle aspiration

25˚C for 5 min.

Synthesis: Incubate at 50˚C for 1 hr.

Inactivation: 70˚C for 15 min.

Removal of RNA template: add 1 µl RNase H (5 U/µl, New England Biolabs)

Heat to 37˚C for 20 min.

Purify with Zymo-5 column – following manufacturer directions, using **7 volumes** (150 µl) of binding buffer (because it's single-stranded DNA), elute two times with 22 µl EB (Elution Buffer) (into same tube)

NOTE: Zymo-5 columns bind a maximum of 5 µg. If you expect larger yields, use more than one column and scale up second strand synthesis as necessary.

Second strand synthesis:

To previous eluate, add:

1.0 µl Random Primers (3 µg/µl, Invitrogen)

Heat to 95˚C for 2 min, cool down in ice for a couple of minutes and spin down for 20 seconds.

Add:

5.0 µl NEB buffer #2 (as alternative use Klenow exo- buffer (Epicentre))

1.0 µl 10 mM dNTPs

1.5 µl Klenow exo- (10 U/µl, Epicentre) (alternatively use DNA polymerase I)

Synthesis: 37˚C for 45 minutes.

Purify with Zymo-5 DNA column (see NOTE above) and:

Elute two times with 50 µl EB each time onto the same tube and proceed to shearing and adaptor ligation for EST building (or expression profiling) via Illumina sequencing.

Shearing:

Up to 2 µg DNA in 100 microliters in EB or TE (in 0.6 ml tubes, Axygen) – fill all other positions in Bioruptor holder with tubes containing 100 µl water. Before beginning, make sure Bioruptor tank water is 4˚C – bail out and replace with cold water if necessary and add a little crushed ice.

Set the controls to shear 30 sec. on, 30 sec. off for 15 cycles

Replace tank water with cold water and a little ice

Repeat shearing, 15 cycles.

(shearing can also be performed with a regular sonicator)

Zymo-5 column concentrate – elute 2 times with 17 µl EB.

End repair:

To eluate (approx. 34 µl), add Epicentre's End-It DNA Repair Kit reagents:

5 µl 10 X buffer

5 µl 1 mM dNTP mix

5 µl ATP

1 µl enzyme mix

Incubate 45 min at room temperature

Clean with Zymo-5 column, elute 2 times with 22 µl EB.

Addition of A overhangs:

To eluate (approx. 44 µl) add:

5 µl 10x Klenow exo- buffer (Epicentre)

1 µl dATP 10 mM

1.5 µl 10 U/µl Klenow Exo- (Epicentre)

Incubate 30 min at 37 C

Clean with Zymo-5 column, elute 2 times with 21.5 µl EB

Adapter Ligation:

Add to eluate from previous step (approx. 42 µl) add:

5 µl 10x NEB buffer 2

1 µl 25 mM ATP (Epicentre)

1 µl 10 µM T-overhang PE adapter (Illumina adapter)

0.5 µl T4 DNA ligase (10 U/µl, Epicentre)

Ligate 1 hour at room temperature or up to overnight in the refrigerator (prefered).

Alternative A. – Zymo column purify, elute 2 times with 6 µl elution buffer each time and proceed to amplification before size fractionation. Only do this if there is a small amount of material. It is generally better to size-fractionate before amplification.

Alternative B. – Zymo column purify, elute 2 times with 10 µl elution buffer and size fractionate (see below)

Size fractionation:

Run de DNA in a 2.5 % agarose gel with 100bp size standard.

Cut out and retain the 200 – 500bp fraction, carefully avoiding any unincorporated adapter. If desired, also cut the 500-700 bp fraction. These two fractions should constitute most of the sheared cDNA. The 200-500 bp size range may be less biased against short transcripts. Recover DNA with Qiagen gel extraction kit and MinElute columns. Dissolve gel at room temperature. Follow

manufacturer instructions, being sure to let wash buffer stand on column for at least 5 minutes before spinning through. Elute 2 times with 15 µl EB.

Library amplification

Quantify the DNA concentration using Qubit (Invitrogen) or some other high resolution fluorometer. Use 25-100 ng of template in a 12x cycle amplification using the Illumina PCR primers.

To DNA template add:

10 µl 5x PCR buffer HF (New England Biolabs)

1 µl 10 µM dNTP

1 µl 10 µM PCR primer mix (PE primer 1+2)

0.5 µl Phusion DNA polymerase (2 U/µl, NEB)

water to 50 µl

PCR conditions:

step1- 98C for 30 secs

step2- 98C for 10 secs

step3- 65C for 30 secs

step4- 72C for 30sec

step5- go to step 2 x12 times

step6- 72C for 5 minutes

step7- 10C hold

Clean with Zymo-5 column, elute 2 times with 12 µl EB

Size fractionation:

Run the PCR product in a 2 % agarose gel with 100bp size standards.

Cut the 300 – 600bp amplified fraction, carefully avoiding any unincorporated PCR primer. The PCR product will run at approximately 100bp higher size than the DNA template used in the PCR.

Gel purify using MinElute columns (QIAGEN) and elute in EB. The volume of EB will depend on the intensity of the DNA band. In general 20 µl will be sufficient. Add 2 µl of 1% Tween-20 before storing in the freezer.

Quantify the DNA concentration using Qubit (Invitrogen) or some other high resolution fluorometer and dilute a DNA sample to the recommended concentration for Illumina sequencing (5 or 10 nmolar).

PE Adapter

T-overhang PE Adapter: 10 µ*M.* A modified Solexa© adapter (2006 Illumina, Inc., all rights reserved). Prepare 10 µ*M* stocks of PE adapters in 1X NEB buffer #2 (or buffer with a 50mM final concentration of NaCl). Anneal adapters in PCR machine by going from 96C to 25C by decreasing the temperature 1C per minute.

Below, example PE adapter sequences with no barcode (PEnoBC). Asterisk denotes a phosphorothioate bond introduced to confer nuclease resistance to the double-stranded oligo, "/5Phos/" denotes a phosphate group.

PEnoBC-top:

5´- ACACTCTTTCCCTACACGACGCTCTTCCGATC*T -3´

PEnoBC- bottom:

5´- /5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG -3´

Library Amplification

Modified Solexa© Amplification primer mix (2006 Illumina, Inc., all rights reserved): 10 μ*M.*

PEprimer1

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

PEprimer2

5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'

References

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One **3:** e3376.

CATCHEN, J. M., A. AMORES, P. A. HOHENLOHE, W. A. CRESKO and J. H. POSTLETHWAIT, 2011 Stacks: de novo genotype analyses from short-read sequence data. Genes, Genomes and Genetics.

HOHENLOHE, P. A., S. BASSHAM, P. D. ETTER, N. STIFFLER, E. A. JOHNSON *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet **6:** e1000862.

MILLER, M. R., T. S. ATWOOD, B. F. EAMES, J. K. EBERHART, Y. L. YAN *et al.*, 2007a RAD marker microarrays enable rapid mapping of zebrafish mutations. Genome Biol **8:** R105.

MILLER, M. R., J. P. DUNHAM, A. AMORES, W. A. CRESKO and E. A. JOHNSON, 2007b Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res **17:** 240-248.

ZERBINO, D. R., and E. BIRNEY, 2008 Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:** 821-829.

Table S1 Sequences per fish

Fish	Sequences	Stacks	SNPs
male	4,694,285	82,143	10,100
female	4,231,941	81,503	9,447
progeny_1	643,297	57,233	4,663
progeny_2	1,789,897	65,364	9,484
progeny_3	2,715,537	67,077	10,178
progeny_4	1,784,627	65,837	9,618
progeny_5	2,134,445	90,791	9,104
progeny_6	2,084,585	74,669	8,809
progeny_7	3,346,939	84,837	10,616
progeny_8	2,836,206	76,847	10,052
progeny_9	3,372,173	72,823	10,315
progeny_10	2,380,042	69,991	9,751
progeny_11	2,383,122	76,977	9,574
progeny_12	3,142,402	87,409	10,176
progeny_13	1,837,842	80,711	8,420
progeny_14	2,194,250	94,163	8,383
progeny_15	744,441	59,356	5,003
progeny_16	3,925,884	433,923	12,626
progeny_17	3,051,835	393,452	10,804
progeny_18	3,141,564	394,713	11,129
progeny_19	2,379,993	296,140	8,803
progeny_20	3,300,759	397,593	11,028
progeny_21	3,674,007	387,798	11,676
progeny_22	1,994,292	257,425	7,895
progeny_23	2,812,816	343,957	9,967
progeny_24	2,743,864	331,489	9,902
progeny_25	2,660,642	317,927	9,298
progeny_26	3,960,192	375,931	11,480
progeny_27	3,762,648	387,889	11,190
progeny_28	2,505,011	314,252	9,321
progeny_29	2,668,092	320,277	9,342
progeny_30	2,038,211	278,254	7,501
progeny_31	1,901,964	248,975	7,372
progeny_32	2,336,896	303,702	8,719
progeny_33	4,124,569	436,789	12,439
progeny_34	2,940,943	312,995	10,008
progeny_35	2,283,742	236,572	8,711

progeny_36	2,358,741	301,974	8,575
progeny_37	2,685,887	297,729	9,838
progeny_38	2,300,697	286,404	8,847
progeny_39	4,050,016	401,603	11,855
progeny_40	2,359,506	312,359	8,253

'Sequences' lists the number of retained reads. 'Stacks' lists the number of different stacks formed by the software for each individual. 'SNPs' lists the number of single nucleotide polymorphisms found in the RAD-tags for each animal.

Table S2 Data for calculating conserved syntenies among gar (Loc), human (Hsa), zebrafish (Dre), and stickleback (Gac).

	Hsa	Loc	Dre	Gac
Hsa	0	0.1589205	0.3408696	0.310559
Loc	0.1589205	0	0.24	0.2492163
Dre	0.3408696	0.24	0	0.1942857
Gac	0.310559	0.2492163	0.1942857	0

Dre vs Loc

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	LG	Count	Probability
1	0	1	1	11	0	2	0	2	0	0	0	4	2	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	9	0.3103448
2	1	0	0	0	4	2	0	0	11	10	0	0	2	2	2	0	0	0	8	0	0	0	0	0	0	0	0	0	0	9	0.3103448
3	1	0	0	0	3	5	0	3	0	13	0	4	11	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0.2758621
4	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.0344828
5	2	10	1	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	11	12	3	0	0	0	0	0	1	0	9	0.3103448
6	0	1	0	1	1	0	0	0	0	6	0	2	2	4	1	1	3	5	0	0	0	0	0	0	0	0	0	0	0	11	0.3793103
7	0	2	5	6	3	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	2	0	0	0	8	3	0	10	0.3448276
8	11	2	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0	2	2	5	3	0	0	0	7	0	0	0	0	9	0.3103448
9	0	0	0	1	0	0	0	0	1	0	0	14	0	2	3	0	4	0	0	0	0	0	0	0	0	0	0	0	0	6	0.2068966
10	2	9	5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	3	0	0	0	0	0	0	0	7	0.2413793
11	0	0	2	0	6	0	0	0	0	0	0	0	0	3	0	0	0	2	3	0	0	0	0	0	2	0	0	0	0	6	0.2068966
12	0	0	0	0	3	0	0	0	1	6	0	1	5	0	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0.2413793
13	1	0	0	2	17	0	4	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
14	0	0	0	2	0	11	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	5	0.1724138
15	0	5	1	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	10	0	0	0	3	0	0	0	6	0.2068966
16	1	0	0	0	0	3	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	11	0	1	0	0	0	5	0.1724138
17	5	0	0	0	2	3	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0.2068966
18	0	3	3	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	0	0	0	1	5	1	0	3	2	0	0	9	0.3103448
19	0	0	0	0	1	8	0	0	0	0	12	0	0	0	1	1	0	0	0	0	0	0	0	8	0	0	0	0	1	7	0.2413793
20	7	0	0	3	0	0	2	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
21	1	0	3	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0	0	0	0	2	0	6	0.2068966
22	1	0	0	1	6	0	0	1	0	0	0	3	0	1	0	4	0	0	4	0	0	0	0	0	1	0	0	0	0	9	0.3103448
23	3	0	3	8	1	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	9	0	0	0	0	6	0.2068966
24	0	0	2	0	0	2	0	1	7	0	0	0	6	1	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	8	0.2758621
25	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	7	0	0	0	4	0	0	5	0.1724138
	36	33	26	37	51	42	18	21	23	36	23	28	28	17	25	22	11	17	17	20	23	19	7	20	19	7	10	7	1		

Avg Prob

0.24

Hsa vs Loc

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	LG	Count	Probability
1	1	0	3	0	0	17	0	0	0	11	0	0	0	0	0	11	1	1	0	0	0	1	0	11	19	0	0	1	0	11	0.3793103
2	6	1	0	6	0	0	0	0	1	0	0	12	0	6	0	7	1	0	0	0	0	0	1	0	0	0	0	0	0	9	0.3103448
3	0	0	2	0	14	0	0	0	0	0	2	3	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
4	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.0344828
5	0	8	0	0	0	15	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
6	10	0	0	0	0	0	0	1	3	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
7	0	1	0	0	0	0	0	5	2	0	5	0	5	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	7	0.2413793
8	5	0	0	0	0	0	0	0	5	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
9	0	5	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	3	0.1034483
10	0	0	0	1	18	0	0	2	3	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
11	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6	12	5	0	6	0.2068966
12	0	0	0	10	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	3	0.1034483
13	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
14	0	0	0	0	7	0	8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
15	0	0	8	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	4	0.137931
16	1	0	0	0	2	0	2	0	0	0	0	0	14	0	0	0	0	0	0	0	0	1	12	0	0	0	0	0	0	6	0.2068966
17	0	7	0	0	0	0	0	0	0	14	1	0	0	0	17	0	0	0	0	0	0	8	0	0	0	1	0	0	0	6	0.2068966
18	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.0689655
19	0	8	0	0	0	7	1	1	0	3	0	0	0	0	0	0	0	0	10	0	0	0	0	1	0	0	0	0	0	7	0.2413793
20	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	12	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
21	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0.0689655
22	0	0	0	0	2	0	0	2	0	0	0	9	1	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	5	0.1724138
X	5	0	2	0	0	0	4	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	4	0.137931
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	28	30	23	32	43	39	16	20	21	28	16	24	20	16	19	21	9	13	10	18	20	15	14	13	19	7	12	6	0		

Avg Prob

0.1589205

Gac vs Loc

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	LG	Count	Probability
I	0	10	4	0	1	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	10	1	1	0	4	0	0	0	9	0.3103448
II	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	10	0	0	5	0.1724138
III	0	0	0	1	3	0	0	0	14	10	0	0	0	1	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	7	0.2413793
IV	0	0	0	4	0	11	5	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0.137931
V	1	0	0	0	4	0	0	0	0	4	0	1	5	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0.2068966
VI	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	11	2	0	0	0	0	0	0	0	0	0	0	0	0	3	0.1034483
VII	0	7	5	5	4	3	2	1	1	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	2	0	6	0	12	0.4137931
VIII	0	0	0	3	0	0	0	0	2	8	0	0	0	5	0	0	2	0	9	0	0	0	0	0	0	0	0	0	0	6	0.2068966
IX	0	0	0	10	2	7	0	2	0	4	0	3	4	0	2	7	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0.3103448
X	0	0	0	0	0	7	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	3	0.1034483
XI	1	0	0	0	2	6	0	2	0	8	0	7	9	0	12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	9	0.3103448
XII	3	1	4	8	2	0	0	1	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	17	0	0	0	0	8	0.2758621
XIII	10	7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	6	0	0	0	0	0	0	0	0	5	0.1724138
XIV	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	13	0	0	0	0	0	0	0	0	4	0.137931
XV	6	0	0	3	0	4	9	0	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	9	0.3103448
XVI	0	0	0	1	0	0	1	0	0	0	0	13	0	2	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	5	0.1724138
XVII	0	0	1	2	7	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	1	0	0	0	0	5	0.1724138
XVIII	13	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.0689655
XIX	1	0	4	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	0	0	0	4	0	0	6	0.2068966
XX	1	0	0	0	0	5	0	0	0	0	11	0	0	0	0	0	0	0	0	1	0	1	0	7	0	2	0	1	0	8	0.2758621
XXI	0	0	1	0	1	0	0	0	8	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	6	0.2068966
	38	31	24	37	41	43	19	29	25	35	23	26	18	12	23	19	10	13	14	18	20	16	13	13	18	8	14	7	0	28	0.9655172

Avg Prob

0.2492163

Dre vs Hsa

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	LG	Count	Probability
1	1	4	0	6	0	0	1	0	1	1	0	0	2	0	0	1	0	0	1	1	2	2	0	0	12	0.5217391
2	7	0	2	0	1	2	1	3	0	3	0	0	0	4	0	0	1	2	6	0	0	0	0	0	11	0.4782609
3	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	8	19	0	5	0	0	6	0	0	5	0.2173913
4	0	0	0	0	0	0	1	0	0	1	0	7	0	0	0	0	0	0	0	0	0	1	0	0	4	0.173913
5	0	1	0	0	2	1	1	1	11	0	1	6	1	0	0	2	3	0	0	0	0	3	1	0	13	0.5652174
6	4	4	3	0	0	0	0	0	0	0	0	1	0	0	1	2	3	0	1	5	0	1	1	0	11	0.4782609
7	0	3	0	0	0	0	0	0	2	0	11	1	0	1	4	2	3	0	0	0	0	0	0	0	8	0.3478261
8	4	2	2	0	1	0	0	2	3	0	0	3	0	0	0	0	0	0	2	1	0	2	2	0	11	0.4782609
9	1	14	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	5	0.2173913
10	0	1	0	0	3	0	2	1	4	0	0	1	1	0	0	0	3	0	0	0	3	2	0	0	10	0.4347826
11	2	0	8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	1	0	0	0	0	5	0.2173913
12	1	1	0	0	0	0	3	1	0	3	0	0	0	0	0	3	7	0	0	0	0	1	0	0	8	0.3478261
13	4	4	0	1	0	1	0	0	0	13	0	0	0	5	0	0	0	0	0	0	0	0	0	0	6	0.2608696
14	1	1	0	1	10	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	4	0	7	0.3043478
15	0	2	0	0	0	0	2	0	0	0	3	0	0	0	0	0	5	0	3	0	1	0	0	0	6	0.2608696
16	9	0	2	0	0	1	2	3	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	7	0.3043478
17	4	2	0	0	0	1	0	2	0	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0	7	0.3043478
18	0	1	1	0	0	0	1	0	0	0	7	0	0	0	3	3	0	0	3	0	0	1	0	0	8	0.3478261
19	11	0	2	0	0	2	4	2	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	8	0.3478261
20	2	0	0	2	0	7	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	5	0.2173913
21	1	0	0	0	4	0	0	0	6	0	2	0	0	0	1	0	0	0	0	0	0	0	2	0	6	0.2608696
22	3	2	4	0	0	1	0	0	0	2	0	1	0	0	0	1	0	0	4	0	0	1	0	0	9	0.3913043
23	10	0	1	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	2	0	6	0.2608696
24	1	0	2	0	2	0	2	1	0	1	0	0	0	1	0	2	1	1	0	0	0	0	1	0	11	0.4782609
25	0	0	0	0	0	0	1	0	0	1	3	2	0	0	1	6	0	0	0	0	0	1	0	0	7	0.3043478
	66	42	29	10	23	17	25	16	27	26	27	27	7	15	11	30	48	5	28	13	6	21	14	0		

Avg Prob

0.3408696

Gac vs Hsa

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	LG	Count	Probability
I	0	4	1	0	0	0	3	0	0	0	7	0	1	0	0	0	5	0	6	0	1	0	0	0	8	0.3478261
II	0	1	0	0	0	0	0	0	0	0	8	0	1	0	5	3	0	0	0	0	0	0	0	0	5	0.2173913
III	7	0	1	0	0	2	1	2	0	2	0	0	0	3	0	0	0	3	4	0	0	0	0	0	9	0.3913043
IV	0	1	0	2	10	0	1	0	0	1	0	5	0	0	0	0	0	0	0	0	0	1	4	0	8	0.3478261
V	0	0	0	0	0	0	2	0	0	4	0	0	0	0	0	5	9	0	0	0	0	1	0	0	5	0.2173913
VI	4	3	0	0	0	1	0	0	0	11	0	0	1	1	0	0	0	0	0	0	1	0	0	0	7	0.3043478
VII	1	2	0	0	3	0	1	0	2	0	8	1	1	1	0	2	6	0	0	0	2	0	1	0	13	0.5652174
VIII	3	2	2	2	0	2	0	0	0	0	0	0	0	0	1	0	0	0	8	0	0	0	1	0	8	0.3478261
IX	4	1	0	5	0	0	1	0	1	3	0	0	0	0	0	3	4	0	2	1	0	3	0	0	11	0.4782609
X	10	0	2	0	0	2	2	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	7	0.3043478
XI	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	7	14	0	5	0	0	7	0	0	6	0.2608696
XII	14	0	2	0	0	1	0	0	1	0	0	4	0	0	0	0	0	0	0	5	0	0	2	0	7	0.3043478
XIII	0	3	0	0	3	0	0	2	7	0	0	7	0	0	0	0	0	0	0	0	0	5	1	0	7	0.3043478
XIV	0	1	0	0	2	0	0	0	14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0.173913
XV	5	3	1	1	0	3	0	1	0	0	0	0	0	7	1	0	0	0	0	0	0	0	0	0	8	0.3478261
XVI	1	12	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	6	0.2608696
XVII	1	0	7	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	3	1	0	0	0	5	0.2173913
XVIII	0	0	0	0	0	6	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	3	0.1304348
XIX	0	0	0	0	0	0	2	0	0	1	4	3	0	0	3	8	0	0	0	0	0	1	1	0	8	0.3478261
XX	9	0	2	0	0	0	4	3	0	0	0	1	0	0	0	0	2	1	1	0	0	0	0	0	8	0.3478261
XXI	1	0	2	0	3	0	1	2	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	7	0.3043478
	60	33	21	10	21	17	21	12	26	23	27	24	5	16	10	28	40	5	27	10	5	18	11	0		

Avg Prob

0.310559