Genomics 95 (2010) 261–267

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

BAC library for the amphipod crustacean, *Parhyale hawaiensis*

Ronald J. Parchem [a,1], Francis Poulin [b,1,2], Andrew B. Stuart [d], Chris T. Amemiya [d], Nipam H. Patel [a,b,c,*]

[a] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3200, USA
[b] Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA
[c] Center for Integrative Genomics, University of California, Berkeley, CA 94720-3200, USA
[d] Genome Resource Center, Benaroya Research Institute, 1201 Ninth Avenue, Seattle, WA 98101, USA

ARTICLE INFO

Article history:
Received 19 August 2009
Accepted 9 March 2010
Available online 16 March 2010

Keywords:
Evolution
Ecdysozoa
Crustacean
Arthropod
Amphipod
BAC library

ABSTRACT

Bacterial artificial chromosomes (BACs) are capable of propagating large fragments of DNA and have become an invaluable tool for studying genome biology. To fill a phylogenetic gap in available genomic sequence and to complement ongoing molecular and genetic studies, we have generated a BAC library for the marine amphipod crustacean, *Parhyale hawaiensis*. The library was generated from genomic DNA isolated from whole adult animals and comprises 129,024 individual clones. The median insert size is ~140 kb and the genomic coverage is approximately five genome equivalents. We have screened the *Parhyale* BAC library for developmentally relevant genes and characterized the genomic organization of four genes: a *hedgehog* ortholog, and three *Pax3/7* paralogs. Preliminary analysis suggests that introns are larger and more prevalent in *Parhyale* than in other arthropods whose genomes have been sequenced, which may partly account for the large genome size in *Parhyale*.

Introduction

Bilaterian animals are separated into two main clades, the deuterostomes and the protostomes. The deuterostomes include the chordates (e.g. humans) and echinoderms (e.g sea urchins), while the protostomes are divided into two large clades, the lophotrochozoans and the ecdysozoans. Molluscs (e.g. snails) and annelids (e.g. earthworms) are part of the lophotochozoans, while nematodes (e.g. *C. elegans*) and arthropods (e.g. insects, crustaceans) are ecdysozoans. Of these groups, the ecdysozoans contain the vast majority of described animal species, due in large part to the diversity of arthropods. Our current understanding of the phylogenetic relationship of some of the major groups within the Ecdysozoa is shown in Fig 1.

Within the Ecdysozoa lie two powerful genetic model systems, *Drosophila melanogaster* and *C. elegans*. These two animals were among the first to have their genome completely sequenced [1,2], and the genomes of several closely related species have been sequenced in order to complement and leverage the large body of work involving these two model species [3,4]. While *Drosophila* and *C. elegans* remain premier genetic systems for studies in many fields of biology, our initial reliance upon these two organisms may have biased some of our views of animal evolution and development.

For example, the genomes of *Drosophila* and *C. elegans* are somewhat unusual in that they are relatively small, which undoubtedly contributed to the decision to sequence them. However, many studies suggest that these two species may not be representative of typical extant bilaterians. For instance, *WNT* genes are a family of highly conserved cell-cell signaling molecules whose founding member is *wingless*. In *Drosophila* there are seven (7), and in *C. elegans* there are five (5) *WNT* family members [5]. In contrast, the human genome contains nineteen (19) *WNT* genes, which at first glance might suggest an expansion of this family in the vertebrate lineage. However, more recent analyses of *WNT* family members in lophotrochozoans and the phylogenetically basal cnidarian *Nematostella vectensis* has revealed thirteen (13) ancestral *WNT* subfamilies [6–8]. This strongly suggests that the small number of *WNT*s in *Drosophila* and *C. elegans* is due to gene loss. More extensive evidence of gene loss for these two model species was provided by the finding of several important genetic pathways in Cnidaria that are absent in *Drosophila* and *C. elegans* [9]. The question of when these losses occurred in the evolution of Ecdysozoa remains largely unanswered, and it is quite possible that at least some of these losses represent cases of independent loss within the nematode and insect lineages. In addition to gene loss, the genomes of *Drosophila* and *C. elegans* are peculiar in that they appear to have undergone compaction, as evidenced by the loss and shortening of introns, and a general decrease in intergenic distance [10,11].

Understanding when these genomic changes occurred and what role they played in the evolution of extant animals requires genomic

* Corresponding author. University of California at Berkeley, Department of Molecular and Cell Biology, 519A LSA # 3200, Berkeley, California 94720-200, USA. Fax: +1 510 643 5022.
E-mail address: nipam@uclink.berkeley.edu (N.H. Patel).
[1] These authors contributed equally to this work.
[2] Present address: Genzyme Corporation, One The Mountain Road, Framingham, MA 01701, USA.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.03.005

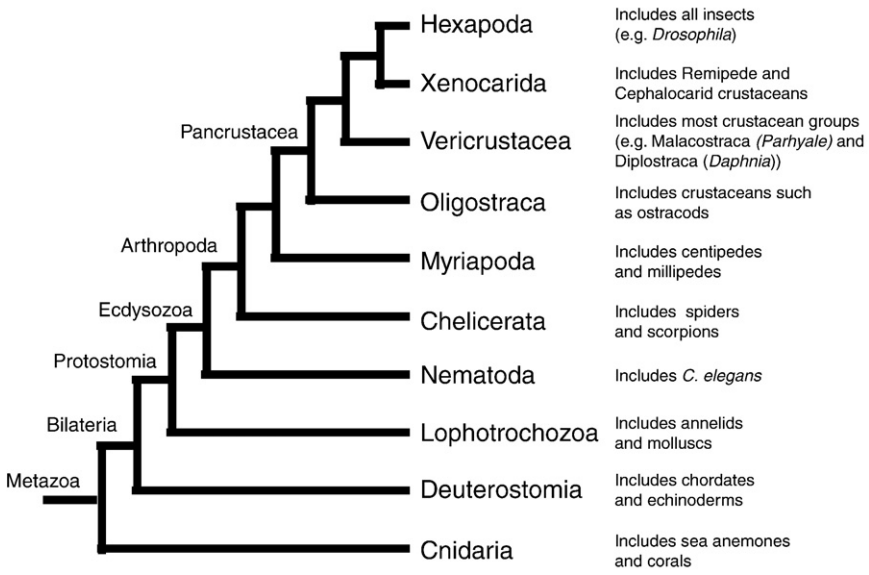

Fig. 1. Evolutionary relationships. Phylogenetic relationships of the major groups of animals within Ecdysozoa. Crustaceans are paraphyletic and *Parhyale* is a member of the Vericrustacea lineage [21]. Tree topology and nomenclature based upon Regier *et al*, 2010 [21].

sequence from a broader set of species. Recent advancements in sequencing and bioinformatics have enabled genomic studies of a more diverse set of animals, without having to focus exclusively on established model species with reduced genome size. Indeed, while the genomes of several ecdysozoans have been sequenced, the vast majority belong to two clades: Insecta, with a heavy bias toward holometabolous insects, especially dipterans [3,12–18], and Nematoda [2,4,19]. Therefore, our understanding of genome evolution within Ecdysozoa will benefit greatly from obtaining genomic sequence from additional species representing a broader taxonomic sampling [20]. As a case in point, the phylogenomic analysis of 2.6 Mb of sequence from 62 single-copy genes of 75 arthropods was required to resolve the deep phylogenetic history of the major arthropod lineages [21].

Since arthropods are not only species-rich, but also morphologically and developmentally diverse, they are critically important for the study of evolution and development (evo-devo). A great deal is known about *Drosophila* development due to its powerful genetic tools and large research community. In comparison we know far less about the development of other arthropods, thus more taxonomic sampling is necessary to help us understand the forces underlying the radiation of this group [21]. To this end, the marine amphipod, *Parhyale hawaiensis*, has been developed as a new model system for evo-devo studies [22–24]. Part of a morphologically diverse group of crustaceans termed malacostracans, *Parhyale* is closely related to economically important animals such as shrimps, crabs, and lobsters, and its phylogenetic position allows important comparisons to reconstruct ancestral features within insects and Pancrustacea [21,25].

In an effort to develop a more powerful genetic system for evo-devo studies, and to begin to address issues of genome evolution within arthropods and, more broadly, in Ecdysozoa, we have generated a bacterial artificial chromosomes (BAC) library for *Parhyale*. BACs are a plasmid vector-based system capable of accommodating large inserts of DNA (>100 kb) [26–28]. Using an F-factor origin of replication, these bacterial plasmids are maintained as a single copy within *E. coli* and allow for faithful propagation of large DNA fragments. These libraries can then be used for molecular and genomic studies such as positional cloning [29] or comparing gene structures and synteny across different species [30]. Moreover, BACs are a convenient tool for the initial physical and genetic mapping of genomic regions without whole genome sequencing. Thus, BAC libraries are an excellent means of studying genome level questions, particularly for those interested in comparative studies of specific genomic regions. Highlighting their usefulness, BAC libraries have now been constructed for hundreds of species ranging from bacteria to plants to animals. Here we report the building and initial characterization of a *Parhyale hawaiensis* BAC library.

Results and discussion

We have constructed a BAC library for the amphipod crustacean, *Parhyale hawaiensis,* that consists of 129,024 clones with an average insert size of 140 kb. The genome size of *P. hawaiensis* is estimated to be 3.6 Gb (N.H.P.and Aziz Aboobaker, unpublished data), and therefore the coverage of our library is estimated to be close to five genome equivalents. The *Parhyale* BAC clones were robotically picked, grown, and stored in 384-well plates and they were arrayed onto several sets of nylon filters for screening. To assess the quality of the library, we screened the filters with gene-specific probes derived from cDNAs we cloned from *Parhyale* embryos for developmental and evolutionary studies. Probes were designed outside of conserved domains to minimize their potential for cross-hybridization. The identity of BAC clones identified by screening was confirmed by PCR and/or Southern analysis. Table 1 summarizes the results of screening the *Parhyale* BAC library with probes for twenty (20) single-copy genes. This initial study confirmed the anticipated 5X coverage of the library. Based on the success of our initial screening, we identified an additional 50 BACs and analyzed their insert size by pulse-field gel electrophoresis. This analysis revealed the average insert size to be 140 kb (Fig. 2), with inserts ranging in size from 95 kb to 170 kb. The majority of clones (46 of 70) have inserts of 130 to 150 kb (Fig. 3). Importantly, our results demonstrate the quality of this BAC library, which contains inserts of expected size, and possesses enough genomic coverage to identify several BACs spanning a genomic region of interest.

One reason for investigating the genome of *Parhyale* is that it is larger than that of the classic ecdysozoan models, *Drosophila* and *C. elegans*. These organisms, potentially as a consequence of their rapid life cycle, appear to have undergone a reduction of their genome size resulting in intron loss and smaller gene size. To address this issue specifically, we have used BAC mapping and sequencing to verify whether genes are larger and introns more prevalent in *Parhyale*. Sequence analysis of four different BACs provides preliminary evidence that introns have both greater size and number in *Parhyale* when compared to their ortholog(s) in other species.

Table 1

Gene	Number of Clones
Ph-pax3/7-1	5
Ph-pax3/7-2	6
Ph-pax3/7-3	5
Ph-sloppy-paired-1	8
Ph-hes-1	7
Ph-hes-2	3
Ph-hes-3	7
Ph-hes-4	5
Ph-runt-1	3
Ph-runt-2	1
Ph-engrailed-1	8
Ph-engrailed-2	5
Ph-odd-skipped-1	4
Ph-odd-skipped-2	2
Ph-odd-skipped-3	4
Ph-odd-skipped-4	4
Ph-odd-skipped-5	2
Ph-odd-paired-1	5
Ph-odd-paired-2	4
*Ph-hedgehog**	12* mean = 5

* The probe for *hedgehog* spanned three introns (>100 kb of genomic sequence). This may have caused the number of clones to be higher than the other genes which were screened for using probes made against putative conserved exons.

For example, comparative analysis of the *hedgehog* locus illustrates the differences between *Parhyale* and other metazoans (Fig. 4). In the case of *Drosophila,* the *hedgehog* gene contains two introns that flank a highly conserved 262 bp exon. The first intron in *Drosophila* is 7,778 bp whereas its homolog in *Parhyale* is ten times larger at 77,848 bp. The second intron is 3,590 bp in *Drosophila*, while its homolog is 11,927 bp in *Parhyale*. Interestingly, the 262 bp exon, which is conserved in all metazoans we analyzed, is disrupted by an additional phase 0 intron (5,717 bp) in *Parhyale* (Fig. 4). Therefore, at the *hedgehog* locus, *Parhyale* has approximately ten times the number of intronic base pairs when compared to *Drosophila*. For perspective, this means that at least 71% of our 146 kb BAC insert is intronic. The yellow fever mosquito, *Aedes aegypti*, which has undergone genome expansion due to an increased mobilization of transposable elements, has a slightly smaller *hedgehog* locus than *Parhyale* and no additional intron when compared to other metazoans (Fig. 4). Vertebrates, which in general have expanded genomes with large genes and

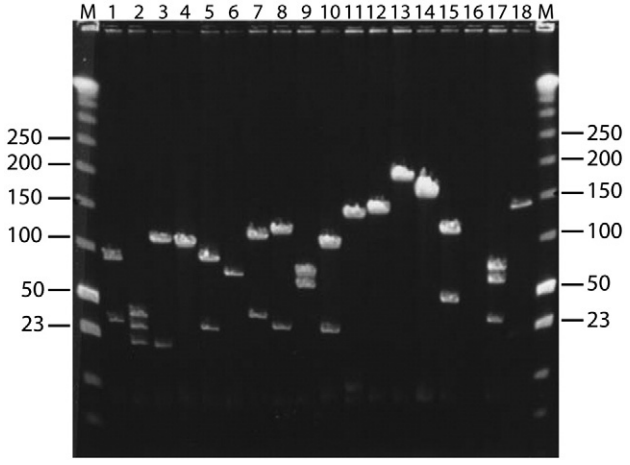

Fig. 2. Sizing and partial characterization of randomly selected recombinants from the *P. hawaiensis* BAC library. Eighteen clones were selected at random from the library before arraying. DNA was isolated using standard miniprep procedure, subjected to *Not*I digestion and analyzed by PFGE as described in *Methods.* Low-range PFGE molecular weight markers (New England Biolabs) are denoted by 'M' and their sizes are given in kb. Miniprep for the BAC clone analyzed in lane 16 did not yield any DNA.

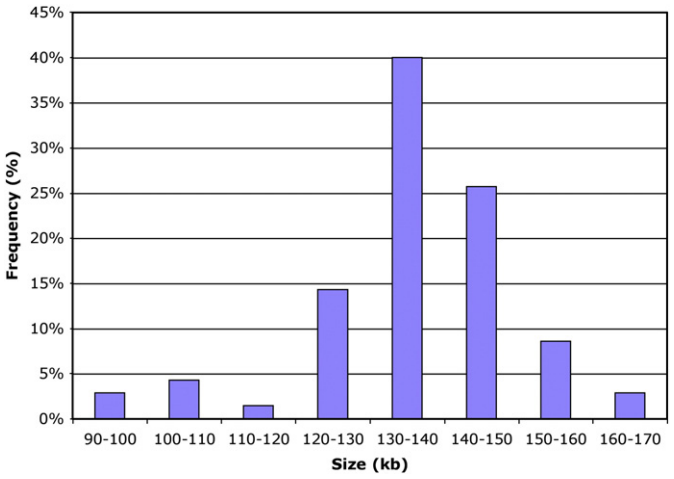

Fig. 3. Size distribution of *Parhyale* BAC clone inserts. A total of 70 clones selected for sequencing were prepared and digested with the restriction enzyme *Not*I. Samples were run on a CHEF gel and insert sizes were calculated manually by comparison with standard size markers.

increased intron number, also have much smaller *hedgehog* introns (Fig. 4).

Analysis of other *Parhyale* BAC sequences suggests similar differences in intron size and number, for example within the Pax group III genes. In arthropods, Pax group III genes are part of the larger Pax family of transcription factors and are homologous to vertebrate Pax3 and Pax7 genes [31–33]. Like other Pax genes, Pax group III are defined by the presence of three domains: the "paired" domain, an octapeptide, and a "paired" class homeodomain [34,35]. In *Drosophila*, they consist of three paralogs; paired (*prd*), gooseberry (*gsb*), and gooseberry-neuro (*gsb-n*), and are important in several developmental processes, namely segmentation and neural development [32,33]. In *Drosophila*, *gsb* and *gsb-n* are syntenic and reside on opposite strands of a 23.5 kb locus. This organization is conserved in *Anopheles* and *Tribolium*, where orthologous genes are found in a 51 kb and 39 kb locus, respectively. The Pax group III genes are not annotated in the *Aedes aegypti* genome, but BLAST searches of the assembly indicate that two homologous genes are possibly syntenic as they are both located in supercontig 1.800, but with the caveat that sequence for the homeodomain is missing from one these genes (http://aaegypti.vectorbase.org). Outside of insects, the branchiopod crustacean, *Daphnia pulex* has undergone an expansion of this family, as there are 6 putative Pax group III genes organized in three clusters of two genes each (F.P. and John Colbourne, unpublished data). The first cluster is 27 kb, the second 63 kb and the third 48 kb, with the six genes ranging from 4 to 18 kb in size. However, not all six *Daphnia* genes are predicted to be functional as two of them have a frameshift or a deletion in the homeodomain sequence.

To better understand the evolution of Pax group III genes in Pancrustacea, we cloned three paralogs from *Parhyale* embryonic cDNA and characterized three BACs containing one gene each (Table 2). The large size of these genes (primary transcript lengths greater than 33 kb, 46 kb, and 18 kb) limits the amount of flanking genomic sequence contained within our BACs, and was not sufficient to uncover any synteny, if it exists, between the *Parhyale* Pax group III genes we cloned. These findings further illustrate the large size of *Parhyale* introns and suggest that intergenic distance may also be greater relative to that in most other sequenced arthropods. Additionally, a comparison of intron number within Pax group III genes across several phylogenetically distant species suggests that *Parhyale* has more introns per gene than currently sequenced insects and shares similarities in this regard to *Daphnia* (Table 2). For example, the three Pax group III genes in *Parhyale* have either five or six introns, whereas insects such as *Drosophila*, *Tribolium* and *Anopheles* possess

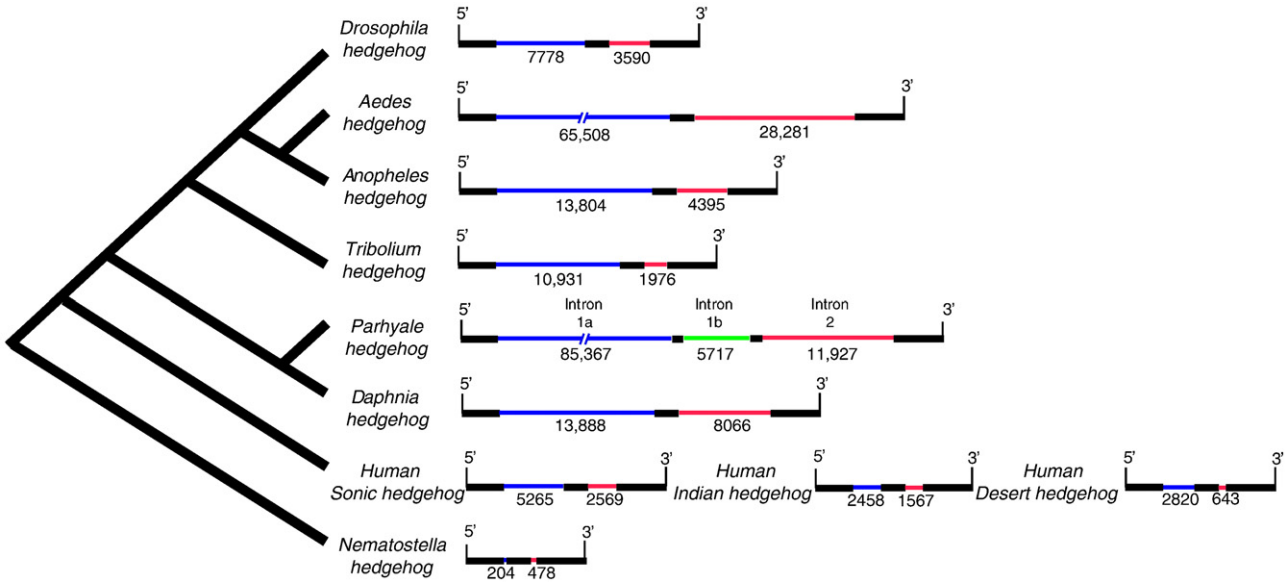

Fig. 4. Comparison of *hedgehog* genomic organization in metazoans. These schematics correspond to the protein-coding region of the *hedgehog* locus. Black bars represent exons (not to scale), the green bar represents an intron in *Parhyale hedgehog* with no corresponding intron in other metazoans in this analysis. Blue and red bars represent homologous introns between species. The large size of intron one in *Parhyale* and *Aedes hedgehog* is denoted by a broken line (and thus are not to scale).

far fewer (mode: 1, range: 1-5). The similarity in intron number between *Parhyale* and *Daphnia* supports the idea that some insects, especially dipterans, may have undergone genome compaction, possibly due to generalized intron loss. Exploring metazoan genomes outside Ecdysozoa also indicates that the *Pax3/7* genes have more introns in the vertebrate lineage. For instance, in *Nematostella* there are four paralogous *PaxD* genes which have three introns each and relatively small size (4 to 8 kb; Table 2). This is in contrast to humans, where *Pax3* and *Pax7* have seven introns, and much larger gene sizes (95 and 103 kb, respectively). To what degree greater intron number in this gene family represents parallel intron gain will require additional taxonomic sampling of genes from the *Pax3/7* family.

Table 2

Species	Gene	Transcript length (bp)	Number of introns	Intron sizes (bp)	Intergenic distance (bp)
Drosophila	*prd*	2,216	1	374	
	gsb	2,315	1	1,031	10,138
	gsb-n	11,062	4	9,712	
Aedes	*pax3/7-1*	>138000	>2	>137000	ND
	pax3/7-2	ND	1?	ND	ND
	pax3/7-3	>69500	>3	>69000	ND
Anopheles	*gsb*	680	1	4,110	32,254
	gsb-n	18,130	3	17,388	
Tribolium	*prd*	28,747	1	27,503	
	gsb	11,004	5	9,771	6,121
	gsb-n	22,316	4	21,299	
Parhyale	*pax3/7-1*	33,721	5	32,257	ND
	pax3/7-2	46,023	6	44,568	ND
	pax3/7-3	18,680	5	17,391	ND
Daphnia	*pax3/7-C*	7,488	6	6,153	51,752
	pax3/7 D	4,231	4	3,433	
	pax3/7 A	4,830	3?	3,052	10,604
	pax3/7 B	11,652	4	10,350	
	pax3/7 E	5,905	5	4,321	24,319
	pax3/7 F	17,997	5	16,506	
Human	*Pax3*	97,268	7	95,251	
	Pax7	105,132	7	102,861	
Nematostella	*PaxD2*	7,650	3	6,555	~10,000
	PaxD1	8,608	3	7,612	
	PaxD3	ND	ND	ND	ND
	PaxD4	5,059	3	4,372	ND

ND = Not determined.
Light grey highlights two potential pseudogenes in *Daphnia*.
Dark grey boxes indicate that no synteny was observed after the analysis of completed genomic sequences.

Interestingly, the large genome size and gene structure in *Parhyale* are reminiscent of the mosquito, *Aedes aegypti*. Analysis of *Aedes* genome indicates that it has increased in size relative to other dipterans due to higher transposable element (TE) content [17]. Consistent with TE mobilization, intron size and intergenic distance have increased, but not intron number [17]. In contrast, our analysis of four BACs suggests that *Parhyale* not only has larger introns and greater intergenic distance, but may also have a global increase in the number of introns per gene (Fig. 4 and Table 2). Unfortunately, the phylogenetic distance between *Parhyale* and other well-characterized genomes, and the limited amount of genomic sequence obtained thus far, did not allow us to characterize the repeat content of *Parhyale*. However, similitudes with *Aedes* suggest that the large genome size of *Parhyale* may be attributable, at least in part, to an increase in the number of repetitive elements. This question is critical to our understanding of the forces that shaped the crustacean genome and can be investigated further when greater sequence data are available. Although preliminary, our survey of *Parhyale* BAC sequences suggests that it may have undergone a unique form of genome expansion that could include an increase in intron number, a phenomenon yet to be fully characterized. It is intriguing that a similar phenomenon of intron gain has recently been described in extant populations of another crustacean, *Daphnia pulex* [36]. A similar process occurring in *Parhyale* could explain the variation in intron position that is observed in Pax group III genes relative to that in other arthropods (Table 2). As an alternative to intron gain, *Parhyale* could represent an arthropod genome possessing ancestral introns that were lost in other arthropod lineages with sequenced genomes. These two alternatives are not mutually exclusive, although intron loss is expected to be more prevalent and less deleterious to organismal fitness relative to intron gain. Further analysis of arthropod

genomes, especially those from crustaceans of varied lineages, and with genomes comparable in size to *Parhyale*, may shed light on this issue.

Practically speaking, the presence of large and frequent introns in *Parhyale* will necessitate careful mapping of BAC inserts and/or the use of multiple BACs to study intron-containing loci. For example, the coding region of *Parhyale hedgehog* occupies a genomic region of 104,631 bp, whereas *Drosophila* only occupies 12,782 bp. If one considers that the *hedgehog* BAC chosen for sequencing was 146 kb, of which 104 kb spanned the coding region, it is apparent that our ability to identify a BAC containing both coding and large upstream or downstream regions is limited. Nevertheless, by "walking" through the BAC library we have been able to analyze very large loci, such as the Hox gene complex of *Parhyale* (N.H.P. and Julia Serano, unpublished data).

In summary, we have generated a BAC library for the emerging model system, *Parhyale hawaiensis*. Our initial characterization of this library indicates that its genomic coverage is 5X and that the average insert size is 140 kb, thereby providing a valuable resource for genomic studies in *Parhyale*. Additionally, we have isolated 70 BAC clones for developmentally regulated genes and have undertaken their complete sequencing. Interestingly, the data presented here suggests significant differences in genome structure, particularly in intron size and frequency, in *Parhyale* relative to the other arthropods that have been sequenced. Importantly, these preliminary findings allow us to generate initial hypotheses regarding genome size evolution. These hypotheses can be tested by both further analyzing genome organization in *Parhyale* and by increasing the overall phylogenetic diversity of characterized genomes, making certain to avoid a bias towards only small genomes. Since *Parhyale* is phylogenetically distant from current model systems and analysis of its genome may in part help understand ecdysozoan genome evolution, we believe this BAC library represents an important tool for comparative evolutionary studies.

Materials and methods

High molecular weight DNA extraction

We isolated high molecular weight (HMW) DNA from forty adult *Parhyale hawaiensis* (20 males and 20 females) collected from an isofemale population established in 2001 (Iso2). The animals were starved for 6 days, rinsed several times in 0.22 μm filter sterilized sea water (FSW), and treated for 48 hours with antibiotics (Tetracycline 30 μg/mL; 1% PenStrep (Invitrogen)). The specimens were rinsed several times with ice-cold FSW, rinsed twice with ice-cold Homogenization Buffer (HB: 0.35 M sucrose, 0.1 M EDTA, 50 mM Tris pH 8.0), minced grossly with needles, and added to 15 mL of ice cold HB. Nuclei were released with a dounce homogenizer on ice, 30 cycles with the loose pestle and 15 cycles with the tight pestle. The homogenate was filtered through three 100 μm cell strainers (BD Falcon), and then two 70 μm cell strainers (BD Falcon). The suspension was pelleted at 3200×g and 4 °C for 15 min. The nuclei pellet was resuspended in 20 mL ice cold HB, filtered through three successive 70 μm cell strainers (BD Falcon), and pelleted at 3200×g and 4 °C for 15 min. The nuclei were resuspended in 500 μL ice cold HB and incubated at 37 °C for 10 min before mixing with 500 μL of 2% Incert agarose (Lonza) prepared in HB and kept at 37 °C. The suspension was dispensed in 80 μL block molds (Bio-Rad) and placed at 4 °C for 10 min. The blocks were incubated at 37 ° C in cell lysis solution (CLS: 1% LDS, 10 mM Tris pH 8.0, 150 mM EDTA pH 8.0) for 4 days, with daily changes of CLS. The CLS was substituted with Block Storage Solution (BSS: 0.2% N-laurylsarcosine, 2 mM Tris pH 8.0, 140 mM EDTA pH 8.0) by rinsing four times in 2 h. The blocks were stored in BSS at 4 °C until processing.

BAC library construction

For size fragmentation, genomic DNA (1/2 block per reaction) was partially digested at 37 ° C for 2.5 hours in 500 μL reaction volumes containing 1 unit of the restriction enzyme *Eco*RI (New England Biolabs) and 100 units *Eco*RI methylase. The quantity of EcoRI and EcoRI methylase was derived from a titration using various amounts of enzymes on 1/6th block per 500 μL reaction. The reaction also contained 2.6 mM spermidine, 0.5 mg/ml BSA (New England Biolabs) and 1x *Eco*RI reaction buffer (0.08 mM S-adenosylmethionine, 2 mM $MgCl_2$, 1 mM DTT). The reaction was stopped by adding 300 μg Proteinase K, 2.9 % *N*-laurylsarcosine, and 0.29 M EDTA, followed by an incubation at room temperature for 1 h. Proteinase K was inactivated by transferring blocks in 15 mL of ½X TE with 15 μL PMSF (100 mM), incubating for 20 min at 4 °C, and repeating 3 times with fresh solution. Blocks were equilibrated in 50 mL ½X TE for 1 hour at 4 °C. DNA fragments were separated on a 1% agarose gel (Pulse-Field certified Bio Rad) using pulse-field gel electrophoresis (BioRad CHEF XA Mapper) in ½X TBE buffer as previously described in Lang *et al* [37]. Gel fragments were excised from the preparative lane that contained HWM DNA, and the HMW DNA was electroeluted at 4 V/cm for 4 hours at 4 °C and then dialyzed at 4 °C for 16 h in 4 L of ½X TE. The sample volume was reduced to ~180 μL by dialyzing at 4 °C against 30% PEG8000 in ½X TE. The CopyControl BAC Cloning Kit (Epicentre) was used for ligation according to the manufacturer's instructions. Approximately 100 ng HMW DNA and 25 ng of the CopyControl pCC1BAC (*Eco*RI) vector were ligated in 50 μL reactions that were incubated overnight at 16 °C. After ligation, reactions were stopped by a brief Proteinase K treatment, deslated by drop dialysis against ddH_2O for 1.5 h using 0.025 μm nitrocellulose filters (Millipore) and the sample volume was reduced to ~20 μL by drop dialysis against PEG8000 (30% in ½X TE) at 4 °C. For transformation, 8 μL of ligation reaction was added to 100 μL electrocompetent cells DH10BT1 (Invitrogen). Electroporations were performed using 2 mm cuvettes and a BTXECM 630 (Harvard Apparatus Inc.) set at 2.5 kV, 225 Ohms, and 25 μF. Transformed cells were recovered in 50 ml conical tubes containing 10 mL SOC (2 transformations per tube), incubated with shaking (250 rpm) for 1 h at 37 °C, and glycerol was added to the transformation mix to a final concentration of 10%. Small aliquots were removed for titration and the remaining transformants were flash-frozen in liquid nitrogen and stored at -80 °C.

Insert size estimation

The EpiLyse Solution (Epicentre) was used to analyze BAC DNA from 36 random white colonies (12 per size fraction), and estimate the frequency and size of inserts. For further analysis, BAC DNA from 18 clones was isolated using a standard alkaline lysis miniprep protocol and a third of each BAC clone preparation was digested using *Not*I (New England Biolabs) and the size of each clone was determined using pulse-field gel electrophoresis (15 h, 1 sec initial time, 20 sec final time, 14 ° C, field angle120°, 6 V/cm) and low range PFG marker (New England Biolabs).

Library arrays and screening

To array the library, transformed cells were grown overnight at 37 °C on LB agar plates (12.5 μg/mL chloramphenicol, 0.4 mM IPTG and 40 μg/mL X-GAL) and single white colonies were transfered into 336×384-well microtiter plates (Genetix) using a colony-picking robot (Norgren Systems). A Total Array System (*Bio*Robotics) was used to replicate the library and prepare high-density nylon filter sets (22 cm×22 cm) containing BAC colony DNA. One filter set was hybridized with α^{32}P-dCTP labeled DNA probes amplified from twenty different cDNAs (Table 1) as previously described [37]. Restriction

digestion of positive clones was performed with *Not*I, followed by pulse-field gel electrophoresis for insert sizing. The *Parhyale* BAC library is stored in the Patel Lab and filters are available upon request.

Acknowledgments

We would like to thank Lisa Chuong for her assistance with BAC data collection and analysis, William Browne for establishing the isogenic *Parhyale* population while in the Patel lab. The U.S. Department of Energy Joint Genome Institute provided sequencing and analyses under the Community Sequencing Program supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and we thank Jane Grimwood for helping to coordinate the sequencing project. F.P. is a Research Fellow of The Terry Fox Foundation through an award from the National Cancer Institute of Canada. N.H.P. is an Investigator of the Howard Hughes Medical Institute.

References

[1] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, et al., The genome sequence of Drosophila melanogaster, Science 287 (2000) 2185–2195.

[2] Genome sequence of the nematode C. elegans: a platform for investigating biology, Science 282 (1998) 2012–2018.

[3] A.G. Clark, M.B. Eisen, D.R. Smith, C.M. Bergman, B. Oliver, T.A. Markow, T.C. Kaufman, M. Kellis, W. Gelbart, V.N. Iyer, D.A. Pollard, T.B. Sackton, A.M. Larracuente, N.D. Singh, J.P. Abad, D.N. Abt, B. Adryan, M. Aguade, H. Akashi, W. W. Anderson, C.F. Aquadro, D.H. Ardell, R. Arguello, C.G. Artieri, D.A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S.A. Bosak, R.K. Bradley, A.D. Brand, M.R. Brent, A.N. Brooks, R.H. Brown, R.K. Butlin, C. Caggese, B.R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S.E. Celniker, J. L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S.W. Clifton, J.M. Comeron, J.C. Costello, J.A. Coyne, J. Daub, R.G. David, A.L. Delcher, K. Delehaunty, C.B. Do, H. Ebling, K. Edwards, T. Eickbush, J.D. Evans, A. Filipski, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A.C. Garcia, A. Gardiner, D.A. Garfield, B.E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravely, A.J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E.A. Gustafson, W. Haerty, M.W. Hahn, D.L. Halligan, A.L. Halpern, G.M. Halter, M.V. Han, A. Heger, L. Hillier, A.S. Hinrichs, I. Holmes, R.A. Hoskins, M.J. Hubisz, D. Hultmark, M.A. Huntley, D.B. Jaffe, S. Jagadeeshan, et al., Evolution of genes and genomes on the Drosophila phylogeny, Nature 450 (2007) 203–218.

[4] L.D. Stein, Z. Bao, D. Blasiar, T. Blumenthal, M.R. Brent, N. Chen, A. Chinwalla, L. Clarke, C. Clee, A. Coghlan, A. Coulson, P. D'Eustachio, D.H. Fitch, L.A. Fulton, R.E. Fulton, S. Griffiths-Jones, T.W. Harris, L.W. Hillier, R. Kamath, P.E. Kuwabara, E.R. Mardis, M.A. Marra, T.L. Miner, P. Minx, J.C. Mullikin, R.W. Plumb, J. Rogers, J.E. Schein, M. Sohrmann, J. Spieth, J.E. Stajich, C. Wei, D. Willey, R.K. Wilson, R. Durbin, R.H. Waterston, The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics, PLoS Biol. 1 (2003) E45.

[5] C.Y. Logan, R. Nusse, The Wnt signaling pathway in development and disease, Annu. Rev. Cell Dev. Biol. 20 (2004) 781–810.

[6] A. Kusserow, K. Pang, C. Sturm, M. Hrouda, J. Lentfer, H.A. Schmidt, U. Technau, A. von Haeseler, B. Hobmayer, M.Q. Martindale, T.W. Holstein, Unexpected complexity of the Wnt gene family in a sea anemone, Nature 433 (2005) 156–160.

[7] B. Prud'homme, N. Lartillot, G. Balavoine, A. Adoutte, M. Vervoort, Phylogenetic analysis of the Wnt gene family. Insights from lophotrochozoan members, Curr. Biol. 12 (2002) 1395.

[8] S.J. Cho, Y. Valles, V.C. Giani, Jr., E.C. Seaver, D.A. Weisblat, Evolutionary dynamics of the Wnt gene family: a lophotrochozoan perspective. Mol. Biol. Evol. (2010), doi:10.1093/molbev/msq052.

[9] N.H. Putnam, M. Srivastava, U. Hellsten, B. Dirks, J. Chapman, A. Salamov, A. Terry, H. Shapiro, E. Lindquist, V.V. Kapitonov, J. Jurka, G. Genikhovich, I.V. Grigoriev, S.M. Lucas, R.E. Steele, J.R. Finnerty, U. Technau, M.Q. Martindale, D.S. Rokhsar, Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization, Science 317 (2007) 86–94.

[10] D.C. Jeffares, T. Mourier, D. Penny, The biology of intron gain and loss, Trends Genet. 22 (2006) 16–22.

[11] Y.H. Loh, S. Brenner, B. Venkatesh, Investigation of loss and gain of introns in the compact genomes of pufferfishes (Fugu and Tetraodon), Mol. Biol. Evol. 25 (2008) 526–535.

[12] S. Richards, R.A. Gibbs, G.M. Weinstock, S.J. Brown, R. Denell, R.W. Beeman, R. Gibbs, G. Bucher, M. Friedrich, C.J. Grimmelikhuijzen, M. Klingler, M. Lorenzen, S. Roth, R. Schroder, D. Tautz, E.M. Zdobnov, D. Muzny, T. Attaway, S. Bell, C.J. Buhay, M.N. Chandrabose, D. Chavez, K.P. Clerk-Blankenburg, A. Cree, M. Dao, C. Davis, J. Chacko, H. Dinh, S. Dugan-Rocha, G. Fowler, T.T. Garner, J. Garnes, A. Gnirke, A. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S.N. Jhangiani, V. Joshi, Z.M. Khan, L. Jackson, C. Kovar, A. Kowis, S. Lee, L.R. Lewis, J. Margolis, M. Morgan, L.V. Nazareth, N. Nguyen, G. Okwuonu, D. Parker, S.J. Ruiz, J. Santibanez, J. Savard, S.E. Scherer, B. Schneider, E. Sodergren, S. Vattahil, D. Villasana, C.S. White, R. Wright, Y. Park, J. Lord, B. Oppert, S. Brown, L. Wang, G. Weinstock, Y. Liu, K. Worley, C.G. Elsik, J.T. Reese, E. Elhaik, G. Landan, D. Graur, P. Arensburger, P. Atkinson, J. Beidler, J.P. Demuth, D.W. Drury, Y.Z. Du, H. Fujiwara, V. Maselli, M. Osanai, H.M. Robertson, Z. Tu, J.J. Wang, S. Wang, H. Song, L. Zhang, D. Werner, M. Stanke, B. Morgenstern, V. Solovyev, P. Kosarev, G. Brown, H.C. Chen, O. Ermolaeva, W. Hlavina, Y. Kapustin, et al., The genome of the model beetle and pest Tribolium castaneum, Nature 452 (2008) 949–955.

[13] J.H. Werren, S. Richards, C.A. Desjardins, O. Niehuis, J. Gadau, J.K. Colbourne, L.W. Beukeboom, C. Desplan, C.G. Elsik, C.J. Grimmelikhuijzen, P. Kitts, J.A. Lynch, T. Murphy, D.C. Oliveira, C.D. Smith, L. van de Zande, K.C. Worley, E.M. Zdobnov, M. Aerts, S. Albert, V.H. Anaya, J.M. Anzola, A.R. Barchuk, S.K. Behura, A.N. Bera, M.R. Berenbaum, R.C. Bertossa, M.M. Bitondi, S.R. Bordenstein, P. Bork, E. Bornberg-Bauer, M. Brunain, G. Cazzamali, L. Chaboub, J. Chacko, D. Chavez, C.P. Childers, J. H. Choi, M.E. Clark, C. Claudianos, R.A. Clinton, A.G. Cree, A.S. Cristino, P.M. Dang, A.C. Darby, D.C. de Graaf, B. Devreese, H.H. Dinh, R. Edwards, N. Elango, E. Elhaik, O. Ermolaeva, J.D. Evans, S. Foret, G.R. Fowler, D. Gerlach, J.D. Gibson, D.G. Gilbert, D. Graur, S. Grunder, D.E. Hagen, Y. Han, F. Hauser, D. Hultmark, H.C.t. Hunter, G.D. Hurst, S.N. Jhangian, H. Jiang, R.M. Johnson, A.K. Jones, T. Junier, T. Kadowaki, A. Kamping, Y. Kapustin, B. Kechavarzi, J. Kim, B. Kiryutin, T. Koevoets, C.L. Kovar, E.V. Kriventseva, R. Kucharski, H. Lee, S.L. Lee, K. Lees, L.R. Lewis, D.W. Loehlin, J.M. Logsdon Jr., J.A. Lopez, R.J. Lozado, D. Maglott, R. Maleszka, A. Mayampurath, D.J. Mazur, M.A. McClure, A.D. Moore, M.B. Morgan, J. Muller, M. C. Munoz-Torres, D.M. Muzny, L.V. Nazareth, et al., Functional and evolutionary insights from the genomes of three parasitoid Nasonia species, Science 327 (2010) 343–348.

[14] Insights into social insects from the genome of the honeybee Apis mellifera, Nature 443 (2006) 931–949.

[15] R.A. Holt, G.M. Subramanian, A. Halpern, G.G. Sutton, R. Charlab, D.R. Nusskern, P. Wincker, A.G. Clark, J.M. Ribeiro, R. Wides, S.L. Salzberg, B. Loftus, M. Yandell, W.H. Majoros, D.B. Rusch, Z. Lai, C.L. Kraft, J.F. Abril, V. Anthouard, P. Arensburger, P.W. Atkinson, H. Baden, V. de Berardinis, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chaturverdi, G.K. Christophides, M.A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C.A. Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M.E. Hillenmeyer, S.L. Hladun, J.R. Hogan, Y.S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J.J. Lin, N.F. Lobo, J.R. Lopez, J.A. Malek, T.C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S.D. Murphy, D.A. O'Brochta, C. Pfannkoch, R. Qi, M.A. Regier, K. Remington, H. Shao, M.V. Sharakhova, C.D. Sitter, J. Shetty, T.J. Smith, R. Strong, J. Sun, D. Thomasova, L.Q. Ton, P. Topalis, Z. Tu, M.F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K.J. Woodford, J.R. Wortman, M. Wu, A. Yao, E.M. Zdobnov, H. Zhang, Q. Zhao, et al., The genome sequence of the malaria mosquito Anopheles gambiae, Science 298 (2002) 129–149.

[16] Q. Xia, Z. Zhou, C. Lu, D. Cheng, F. Dai, B. Li, A draft sequence for the genome of the domesticated silkworm (Bombyx mori), Science 306 (2004) 1937–1940.

[17] V. Nene, J.R. Wortman, D. Lawson, B. Haas, C. Kodira, Z.J. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, Q. Ren, E.M. Zdobnov, N.F. Lobo, K.S. Campbell, S.E. Brown, M. F. Bonaldo, J. Zhu, S.P. Sinkins, D.G. Hogenkamp, P. Amedeo, P. Arensburger, P.W. Atkinson, S. Bidwell, J. Biedler, E. Birney, R.V. Bruggner, J. Costas, M.R. Coy, J. Crabtree, M. Crawford, B. Debruyn, D. Decaprio, K. Eiglmeier, E. Eisenstadt, H. El-Dorry, W.M. Gelbart, S.L. Gomes, M. Hammond, L.I. Hannick, J.R. Hogan, M.H. Holmes, D. Jaffe, J.S. Johnston, R.C. Kennedy, H. Koo, S. Kravitz, E.V. Kriventseva, D. Kulp, K. Labutti, E. Lee, S. Li, D.D. Lovin, C. Mao, E. Mauceli, C.F. Menck, J.R. Miller, P. Montgomery, A. Mori, A.L. Nascimento, H.F. Naveira, C. Nusbaum, S. O'Leary, J. Orvis, M. Pertea, H. Quesneville, K.R. Reidenbach, Y.H. Rogers, C.W. Roth, J.R. Schneider, M. Schatz, M. Shumway, M. Stanke, E.O. Stinson, J.M. Tubio, J.P. Vanzee, S. Verjovski-Almeida, D. Werner, O. White, S. Wyder, Q. Zeng, Q. Zhao, Y. Zhao, C.A. Hill, A.S. Raikhel, M.B. Soares, D.L. Knudson, N.H. Lee, J. Galagan, S.L. Salzberg, I.T. Paulsen, G. Dimopoulos, F.H. Collins, B. Birren, C.M. Fraser-Liggett, D.W. Severson, Genome sequence of Aedes aegypti, a major arbovirus vector, Science 316 (2007) 1718–1723.

[18] Genome Sequence of the Pea Aphid Acyrthosiphon pisum, PLoS Biol. 8 (2010) e1000313.

[19] T.W. Harris, N. Chen, F. Cunningham, M. Tello-Ruiz, I. Antoshechkin, C. Bastiani, T. Bieri, D. Blasiar, K. Bradnam, J. Chan, C.K. Chen, W.J. Chen, P. Davis, E. Kenny, R. Kishore, D. Lawson, R. Lee, H.M. Muller, C. Nakamura, P. Ozersky, A. Petcherski, A. Rogers, A. Sabo, E.M. Schwarz, K. Van Auken, Q. Wang, R. Durbin, J. Spieth, P.W. Sternberg, L.D. Stein, WormBase: a multi-species resource for nematode biology and genomics, Nucleic Acids Res. 32 (2004) D411–D417.

[20] G.E. Budd, M.J. Telford, The origin and evolution of arthropods, Nature 457 (2009) 812–817.

[21] J.C. Regier, J.W. Shultz, A. Zwick, A. Hussey, B. Ball, R. Wetzer, J.W. Martin, C.W. Cunningham, Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences, Nature 463 (2010) 1079–1083.

[22] W.E. Browne, A.L. Price, M. Gerberding, N.H. Patel, Stages of embryonic development in the amphipod crustacean. Parhyale hawaiensis, Genesis 42 (2005) 124–149.
[23] A. Pavlopoulos, M. Averof, Establishing genetic transformation for comparative developmental studies in the crustacean Parhyale hawaiensis, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 7888–7893.
[24] E.J. Rehm, R.L. Hannibal, R.C. Chaw, M.A. Vargas-Vila, N.H. Patel, The crustacean Parhyale hawaiensis: a new model for arthropod development, in: A. Gann, D. Crotty (Eds.), Emerging Model Organisms, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2009, pp. 373–404.
[25] U.W. Hwang, M. Friedrich, D. Tautz, C.J. Park, W. Kim, Mitochondrial protein phylogeny joins myriapods with chelicerates, Nature 413 (2001) 154–157.
[26] K. Osoegawa, P.Y. Woon, B. Zhao, E. Frengen, An improved approach for construction of bacterial artificial chromosome libraries, Genomics 52 (1998) 1–8.
[27] C.T. Amemiya, T.P. Zhong, G.A. Silverman, M.C. Fishman, L.I. Zon, Zebrafish YAC, BAC, and PAC genomic libraries, Methods Cell Biol. 60 (1999) 235–258.
[28] H. Shizuya, B. Birren, U.J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, M. Simon, Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 8794–8797.
[29] K. Ito, K. Kidokoro, H. Sezutsu, J. Nohata, K. Yamamoto, I. Kobayashi, K. Uchino, A. Kalyebi, R. Eguchi, W. Hara, T. Tamura, S. Katsuma, T. Shimada, K. Mita, K. Kadono-Okuda, Deletion of a gene encoding an amino acid transporter in the midgut membrane causes resistance to a Bombyx parvo-like virus, Proc. Natl. Acad. Sci. USA 105 (2008) 7523–7527.
[30] K. Sahara, A. Yoshido, F. Marec, I. Fuková, H. Zhang, C.C. Wu, M. Goldsmith, Y. Yasukochi, Conserved synteny of genes between chromosome 15 of Bombyx mori and a chromosome of Manduca sexta shown by five-color BAC-FISH, Genome 50 (2007) 1061–1065.
[31] M. Buckingham, F. Relaix, The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions, Annu. Rev. Cell. Dev. Biol. 23 (2007) 645–673.
[32] G.K. Davis, J.A. D'Alessio, N.H. Patel, Pax3/7 genes reveal conservation and divergence in the arthropod segmentation hierarchy, Dev. Biol. 285 (2005) 169–184.
[33] G.K. Davis, C.A. Jaramillo, N.H. Patel, Pax group III genes and the evolution of insect pair-rule patterning, Development 128 (2001) 3445–3458.
[34] R. Breitling, J.K. Gerber, Origin of the paired domain, Dev. Genes. Evol. 210 (2000) 644–650.
[35] D.Q. Matus, K. Pang, M. Daly, M.Q. Martindale, Expression of Pax gene family members in the anthozoan cnidarian, Nematostella vectensis, Evol. Dev. 9 (2007) 25–38.
[36] W. Li, A.E. Tucker, W. Sung, W.K. Thomas, M. Lynch, Extensive, recent intron gains in Daphnia populations, Science 326 (2009) 1260–1262.
[37] M. Lang, T. Miyake, I. Braasch, D. Tinnemore, N. Siegel, W. Salzburger, C. Amemiya, A. Meyer, A BAC library of the East African haplochromine cichlid fish Astatotilapia burtoni, J. Exp. Zool. B Mol. Dev. Evol. 306 (2006) 35–44.