Genomewide gene-associated microsatellite markers for the model invasive ascidian, *Ciona intestinalis* species complex

YAPING LIN, YIYONG CHEN, WEI XIONG and AIBIN ZHAN

Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Haidian District, Beijing 100085, China

Abstract

The vase tunicate, *Ciona intestinalis* species complex, has become a good model for ecological and evolutionary studies, especially those focusing on microevolution associated with rapidly changing environments. However, genome-wide genetic markers are still lacking. Here, we characterized a large set of genomewide gene-associated microsatellite markers for *C. intestinalis* spA (= *C. robusta*). Bioinformatic analysis identified 4654 microsatellites from expressed sequence tags (ESTs), 2126 of which successfully assigned to chromosomes were selected for further analysis. Based on the distribution evenness on chromosomes, function annotation and suitability for primer design, we chose 545 candidate microsatellites for further characterization. After amplification validation and variation assessment, 218 loci were polymorphic in at least one of the two populations collected from the coast of Arenys de Mar, Spain ($N = 24$ –48), and Cape Town, South Africa ($N = 24$ –33). The number of alleles, observed heterozygosity and expected heterozygosity ranged from 2 to 11, 0 to 0.833 and 0.021 to 0.818, and from 2 to 10, 0 to 0.879 and 0.031 to 0.845 for the Spanish and African populations, respectively. When all microsatellites were tested for cross-species utility, only 60 loci (25.8%) could be successfully amplified and all loci were polymorphic in *C. intestinalis* spB. A high level of genomewide polymorphism is likely responsible for the low transferability. The large set of microsatellite markers characterized here is expected to provide a useful genomewide resource for ecological and evolutionary studies using *C. intestinalis* as a model.

Keywords: biological invasion, expressed sequence tags, invasive species, microevolution, microsatellite, transferability, tunicate

Received 18 March 2015; revision received 23 September 2015; accepted 21 October 2015

Introduction

In the past three decades, the introduction and spread of nonindigenous species (NIS) at a global scale have dramatically increased, mainly owing to anthropogenic activities such as those associated with global trade (e.g. Carlton & Ruiz 2005; Hulme 2009) and aquaculture (e.g. Lin *et al.* 2015). Numerous NIS have successfully crossed biogeographical barriers formed over millions of years, colonized a variety of new habitats, and become naturalized or even invasive in local environments (e.g. Therriault & Herborg 2008; Zhan *et al.* 2010; Richardson & Rejmánek 2011). Invasive species encounter high selective pressures as they rapidly spread throughout varying environments. The change of environmental conditions during biological invasions can be orders of magnitude higher and/or faster than what species would experience

due to natural processes in their native habitats (Valentine *et al.* 2007; Therriault & Herborg 2008). Consequently, successful invaders offer great opportunities to understand rapid microevolution in response to changing environments in the wild.

The study of microevolutionary processes associated with rapidly changing environments in the wild represents one of the major challenges in ecological and evolutionary studies, especially those aiming to reveal the genetic basis of microevolution using wild populations (Gienapp *et al.* 2008; Hansen *et al.* 2012). Although signatures of microevolutionary processes are clearly visible in many adaptive traits, it is still difficult to link directional or disruptive changes to selection caused by environmental changes, mainly owing to numerous factors such as variable selection across different life stages and generations (e.g. Björklund *et al.* 2009; Collins & de Meaux 2009; Alía *et al.* 2014). In addition, environmental conditions can be complex and variable in time and space, leading to a varying degree of selective pressures

Correspondence: Aibin Zhan, Fax: +86-10-6284-9882; E-mail: zhanaibin@hotmail.com or azhan@rcees.ac.cn

on populations (e.g. Charmantier & Garant 2005). Moreover, different genotypes may have different interactions with environments, making the evaluation of genotype–environment interaction extremely complicated in a context of rapidly changing environments (see review by Procaccini *et al.* 2011). Therefore, the selection of a good model, as well as sound experimental design, is crucial for the study of the genetic basis of rapid microevolution in the wild.

The vase tunicate, *Ciona intestinalis* species complex, represents a good model for the study of microevolution associated with rapidly changing environments in the wild (Procaccini *et al.* 2011; Zhan *et al.* 2015). Phylogenetic and phylogeographical studies have demonstrated that *C. intestinalis* is a species complex consisting of at least four morphologically cryptic but genetically distinct species (spA–D; Caputi *et al.* 2007; Nydam & Harrison 2007; Zhan *et al.* 2010). Further analyses suggested that *C. intestinalis* species complex might have radiated sometime during the late Miocene in the brackish Sarmatian or Pontian seas (Zhan *et al.* 2010). While *C. intestinalis* spC and spD largely remain restricted to their native ranges in the Mediterranean Sea and Black Sea, respectively, highly invasive species including *C. intestinalis* spA (= *C. robusta*; Brunetti *et al.* 2015) and spB have disjunct global distributions (Caputi *et al.* 2007; Nydam & Harrison 2007; Zhan *et al.* 2010). Since the 1930s, these two invasive species have rapidly spread and successfully colonized habitats with dramatically different environmental conditions. For example, they can survive at a wide range of water temperatures from -1 to 35°C and salinities from 12 to 40‰ (Dybern 1967; Carver *et al.* 2003; Theriault & Herborg 2008). In addition, *C. intestinalis* has a small genome (<160 MB), and the whole genome of spA has been sequenced and assembled (Dehal *et al.* 2002; Satou *et al.* 2008). The rapid spread globally, wide distribution in dramatically different environments, and its small genome makes *C. intestinalis* a good model to answer many fundamental questions associated with rapid microevolution at the genome level (Satoh *et al.* 2003; Procaccini *et al.* 2011). However, genomewide genetic markers are still lacking in *C. intestinalis*.

Here, we aimed to develop genomewide microsatellite markers for the model invasive ascidian *C. intestinalis*. Compared to microsatellites derived from genomic resources (i.e. genomic microsatellites), microsatellites mined from expressed sequence tags (ESTs; i.e. EST-microsatellites) have several advantages such as relatively high transferability among related species, a low level of null alleles, and linkage to genes under selection (see review by Ellis & Burke 2007). Consequently, in this study we mined *C. intestinalis* spA ESTs to identify microsatellites and then assigned all microsatellites back

to chromosomes based on the genome assembly version 2.0 at the Joint Genome Institute (JGI) website and the KH assembly of Satou *et al.* (2008). Based on the evenness of distribution of microsatellites on chromosomes, functional annotation and suitability for primer design, we chose a set of 545 microsatellites across the whole genome. We further characterized the polymorphism of the candidate microsatellite loci using two populations: one was collected from a representative location of the likely evolutionary origin of *C. intestinalis* while the other was sampled from the recently invaded range. In addition, we tested the cross-species utility of the characterized microsatellite markers in another invasive species of *C. intestinalis* species complex, *C. intestinalis* spB.

Materials and methods

Data mining, annotation and microsatellite selection

We downloaded all EST sequences of *C. intestinalis* spA from the Joint Genome Institute (JGI) website (<http://genome.jgi.doe.gov/Cioin2/Cioin2.download ftp.html>). All downloaded EST sequences were clustered using CD-HIT-EST (Huang *et al.* 2010) with the sequence identity parameter of 95% to rule out redundant sequences. Nonredundant EST sequences were mined for microsatellites using SCIROKO version 3.4 (Kofler *et al.* 2007) with the parameters of seven repetitions for di-, five repetitions for tri-, and four repetitions for tetra-, and three repetitions for penta- and hexanucleotide repeats.

For putative function determination and annotation, EST-bearing microsatellite sequences were subjected to BLASTX and BLASTN against the NCBI databases. Sequences with sufficient flanking regions and known functions were preferentially selected for mapping on chromosomes of *C. intestinalis*. The well assembled genome (i.e. genome assembly version 2.0, <http://genome.jgi.doe.gov/pages/blast.jsf?db=Cioin2>) and the KH assembly by Satou *et al.* (2008) were employed as the reference for mapping the chromosomal distribution using BLAST. The chromosomal locations of microsatellites were drawn using software MAPCHART version 2.2 (Voorrips 2002). EST-bearing microsatellite sequences successfully assigned to chromosomes were preferentially used in this study. Based on the distribution pattern, especially evenness within and among chromosomes, we chose candidate microsatellites for primer design, validation and polymorphism assessment.

Microsatellite validation and polymorphism assessment

All selected sequences were subjected to PCR primer design using Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, USA): primer length

was between 19 and 25 bp with an expected product size between 100 and 400 bp. All selected microsatellites were amplified at a range of annealing temperatures (45–60 °C) in four individuals and checked in 2% agarose gels. The microsatellites that showed appropriate amplification were further subjected to polymorphism assessment using two populations collected from the coast of Arenys de Mar, Spain ($N = 24$ –48) and Cape Town, South Africa ($N = 24$ –33). Based on our phylogenetic and phylogeographical analyses, the Spanish population is one representative of the likely native range of *C. intestinalis* whereas the South African population represents the recently invaded range (Zhan *et al.* 2010). Genomic DNA was extracted from approximately 10 mg of siphon tissues according to the method of Waters *et al.* (2000). DNA concentration and quality was assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

PCR was carried out using a three-primer system according to Schuelke (2000). All forward primers were 5'-tailed with the M13(-29) forward sequence (5'-CAC-GACGTTGTAAAACGAC-3') and used in combination with an M13 primer of the same sequence but 5'-labelled with one of the four fluorescent dyes: 6-carboxyfluorescein (6-FAM), hexachlorofluorescein (HEX), carboxy-tetramethylrhodamine (TMR) or carboxy-X-rhodamine (ROX). The PCR amplification was carried out in 96-well PCR plates with 12.5 µL reaction volume containing approximately 40 ng of genomic DNA, 1 × PCR buffer, 0.2 mM of each dNTP, 1.5 mM of Mg^{2+} , 0.5 pmol M13-tailed forward primer, 1 pmol reverse primer and 1 pmol fluorescently labelled M13 primer, and 0.25 unit of *Taq* DNA polymerase (Takara Bio Inc.). The cycling PCR profile consisted of an initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, a locus-specific annealing temperature (Appendices S1 and S2, Supporting Information) for 30 s, and 72 °C for 30 s, with a final extension at 72 °C for 5 min. Amplified fragments were separated using an ABI 3730xl automated sequencer (Applied Biosystems, Foster City, CA, USA) with GeneScan™-500 LIZ™ as the internal size standard (Applied Biosystems). The alleles were scored using GeneMapper™ software version 4.0 (Applied Biosystems).

Diversity indices including the number of alleles (A), observed heterozygosity (H_o) and the expected heterozygosity (H_e) were calculated using FSTAT version 2.9.3.2 (Goudet 2001). The presence of null alleles was evaluated using the software MICRO-CHECKER version 2.2.0 (Van Oosterhout *et al.* 2004). The probability of significant deviation from Hardy–Weinberg equilibrium (HWE) was assessed using the Markov chain-based method (Guo & Thompson 1992) implemented in GENEPOP version 3.4 (Raymond & Rousset 1995). Significance criteria were

adjusted for the number of simultaneous tests using the standard Bonferroni correction.

Cross-species utility

All selected microsatellite loci for *C. intestinalis* spA were tested for cross-species utility in *C. intestinalis* spB. The validation and polymorphism assessment procedures were performed using the same protocols as those for *C. intestinalis* spA mentioned above. The polymorphism assessment was based on a population ($N = 24$) collected from the coast of Salzhaff, Germany (Zhan *et al.* 2010).

Results

Data mining and distribution of microsatellites on chromosomes

Among the 684 393 EST sequences deposited at the JGI website, we identified 155 591 nonredundant ESTs after removing redundant sequences using CD-HIT-EST. A total of 4654 (3.0%) microsatellite-containing EST sequences were identified based on the mining criteria from all nonredundant ESTs (Table 1). In general, the occurrence frequency of microsatellites decreased as the motif repeat number increased across all types of motifs (Fig. 1a), and 90.5% of microsatellites had fewer than 12 repeats (Fig. 1b).

The most abundant repeat motif was tri- (48.8%), followed by di- (25.5%) and pentanucleotide (12.6%; Table 1 and Fig. 2). Among all repeat motifs, AAC was the most abundant, accounting for 23.2% and 47.5% of all identified microsatellites and trinucleotides, respectively (Fig. 2). Also, the dinucleotide AT showed a high abundance, accounting for 13.9% and 54.6% of all repeat motifs and dinucleotides, respectively (Fig. 2). By contrast, the GC motif was extremely rare, only accounting for 0.4% of all dinucleotide motifs and only 0.1% of all motifs (Fig. 2).

Table 1 Statistics for microsatellites (SSRs) in expressed sequence tags (ESTs) of *Ciona intestinalis* spA

Motif	No. of SSRs	Percentage (%)	Average repeat number	No. of SSRs per Mbp
Dinucleotide	1188	25.5	13.2	12.6
Trinucleotide	2270	48.8	6.1	24.1
Tetranucleotide	362	7.8	5.2	3.8
Pentanucleotide	586	12.6	3.6	6.2
Hexanucleotide	248	5.3	3.6	2.6
Total	4654	100	–	9.9

Fig. 1 Distribution pattern (a) and ratio (b) with respect to the repeat number of motifs including di-, tri-, tetra-, penta- and hexanucleotide microsatellites derived from the expressed sequence tags (ESTs) of *Ciona intestinalis* sp.A.

For all 4654 microsatellite-containing ESTs, we removed 1211 sequences with short flanking region(s), as these sequences were not suitable for primer design. Afterwards, the remaining 3533 sequences were subjected to chromosomal location assignment based on both the genome assembly version 2.0 at the Joint Genome Institute website (<http://genome.jgi.doe.gov/Cioin2/Cioin2.home.html>) and Satou *et al.*'s (2008) KH assembly available from the Ensembl website (http://asia.ensembl.org/Ciona_intestinalis/Info/Index?db=core). A total of 3043 (86.1%) sequences showed significant hits, 2181 and 862 of which were successfully assigned to chromosomes and scaffolds defined by the genome assembly version 2.0, respectively. For the 2181 sequences successfully mapped on chromosomes, we removed significant clusters by selection of representative sequences in each cluster (25 kb segment as a window for checking clusters). Finally, 1065 candidate microsatellite-containing sequences on chromosomes were selected for further analysis (Appendix S3A, Supporting Information). The number of candidate microsatellites on chromosomes 1–14 was 135, 93, 122, 74, 80, 37, 95, 91, 124, 54, 61, 27 and 72, respectively (Appendices S3A and S4, Supporting Information; chromosome 11 was not included in the genome assembly version 2.0). In general, these microsatellites were relatively evenly distributed on chromosomes (Appendices S3 and S4, Supporting Information). When all 3533 sequences were analysed based on the KH assembly, a total of 3441 sequences (96.8%) showed significant hits, 2126 of which were successfully mapped to chromosomes (Appendix S3B, Sup-

porting Information), and 242, 182, 181, 138, 136, 69, 195, 202, 230, 117, 135, 141, 37 and 121 were relatively evenly distributed on chromosomes 1–14 (Figs 3 and 4).

Microsatellite selection and validation

Based on the distribution pattern on each chromosome, we chose a total of 441 microsatellites that were suitable for primer design. In addition, we randomly chose 104 microsatellite-containing ESTs that were assigned to scaffolds for primer design. The selected 545 candidate microsatellites covered all available 14 chromosomes (except for chromosome 11, Figs 3 and 4; Appendix S3) and all repeat motifs including 112, 381, 45, 1 and 6 loci for di-, tri-, tetra-, penta- and hexanucleotides, respectively.

After validation tests, a total of 432 (79.3%) loci could be amplified successfully, 308 of which showed amplicons <550 bp. For the remaining 124 microsatellites, amplicons were larger than 550 bp, which were likely due to the presence of intron(s) in genomic DNA. Because the amplicons were too large to be genotyped using the 3730xl genetic analyzer, we therefore removed them from further assessment. Further tests on the 308 selected loci showed that PCR amplification of 75 loci was unstable and/or amplified poorly across multiple individuals in the populations surveyed. We discarded these loci from our microsatellite panel, leaving us with 233 loci for further consideration (Fig. 4; Appendix S4, Supporting Information).

Fig. 2 Distribution pattern based on the motif length (a) and the most abundant microsatellites (top five) for each motif type (b, the total number of microsatellites and ratios are shown for each motif in brackets) for *Ciona intestinalis* spA.

Polymorphism assessment and functional annotation

After all 233 selected loci were subjected to polymorphism assessment, 218 loci (93.6%) were polymorphic in at least one of the two populations (Appendix S1, Supporting Information). The average number of alleles/locus was comparable in the two populations,

3.48 alleles/locus for the Spanish population versus 3.49 alleles/locus for the African population. The number of alleles ranged from two to 11 and two to 10 for the Spanish and African populations, respectively. The observed heterozygosity (H_o) varied from 0 to 0.833 for the Spanish population and 0 to 0.879 for the African population, whereas expected heterozygosity (H_e) ranged from 0.021 to 0.818 for the Spanish population and 0.031 to 0.845 for the African population (Appendix S1, Supporting Information). After Bonferroni correction, 44 loci (20.2%) and 32 loci (14.7%) were found to depart significantly from Hardy–Weinberg equilibrium in Spanish and African populations, respectively, 44 and 31 of which exhibited heterozygote deficiency with MICRO-CHECKER tests suggesting null alleles for 44 and 29 of them (Appendix S1, Supporting Information).

After all 218 polymorphic microsatellite-containing ESTs were subjected to BLAST searches against NCBI databases, a total of 193 (88.5%) microsatellite-containing EST sequences were successfully annotated and covered a wide range of genes with various functions (Appendix S1, Supporting Information). In summary, BLASTX/TBLASTX recovered 128 significant matches to the nonredundant protein database, while BLASTN showed significant hits for 65 of the remaining 90 (Appendix S1, Supporting Information).

Cross-species utility

When all 233 microsatellites were subjected to the cross-species utility test in *C. intestinalis* spB, we detected a low level of transferability. Only 60 loci (25.8%) could be successfully amplified and all loci were polymorphic in *C. intestinalis* spB (Appendix S2, Supporting Information). The number of alleles varied from two to 15 alleles per locus, with an average of 4.2 alleles/locus. The

Fig. 3 The total number of microsatellites, markers selected for assessment, and polymorphic markers located on each chromosome of *Ciona intestinalis* spA. Chromosomal location was assessed by BLAST against the KH assembly of Satou *et al.* (2008) (but see Appendix S4, Supporting Information, for the chromosomal location defined based on the genome assembly version 2.0 at the Joint Genome Institute (JGI) website).

Fig. 4 The chromosomal locations of the 177 gene-associated polymorphic microsatellite markers for *Ciona intestinalis* spA. Numbers on the left represent the chromosomal location for each microsatellite marker based on the KH assembly of Satou *et al.* (2008). The names of microsatellite markers were labelled on the right.

observed heterozygosity (H_o) ranged from 0 to 0.864, whereas expected heterozygosity (H_e) ranged from 0.082 to 0.900 (Appendix S2, Supporting Information). A total of 19 loci (31.7%) were found to deviate significantly from Hardy–Weinberg equilibrium after Bonferroni correction for multiple tests, all of which showed heterozygote deficiency and the presence of null alleles based on the MICRO-CHECKER test (Appendix S2, Supporting Information).

Discussion

Owing to the unique evolutionary position between vertebrates and invertebrates and ease of manipulation, *C. intestinalis* has served as a model system in evolutionary and developmental biology for more than one century (see Procaccini *et al.* 2011 and references therein). Recent findings, especially those in invasion biology, have shown that *C. intestinalis* can successfully colonize habitats with dramatically different environmental conditions, such as a wide range of salinities and temperatures (Dybern 1967; Carver *et al.* 2003; Therriault & Herborg 2008). In addition, *C. intestinalis* possess broad tolerance of ecophysiological parameters, such as eutrophic conditions following anthropization (Procaccini *et al.* 2011). All these findings have promoted *C. intestinalis* as a good model for studying the genetic basis of microevolution in rapidly changing environments (Procaccini *et al.* 2011; Zhan *et al.* 2015). To facilitate ecological and evolutionary studies using *C. intestinalis* as a model, we developed a large set of genomewide gene-associated microsatellite markers in this study.

Microsatellite development from ESTs versus genome sequencing data

Genomic resources have increased as sequencing technologies have rapidly developed over the past decade. Many sequencing-based projects, such as those for sequencing transcriptomes and whole genomes, have been completed or are under way in a large number of species. Sequences derived from these projects have proven to be a valuable resource for development of molecular markers such as microsatellites. A large number of microsatellites have been successfully developed by mining data derived from completed sequencing-based projects in numerous species such as turbot *Scophthalmus maximus* (Navajas-Pérez *et al.* 2012) and *Brassica* crops (Shi *et al.* 2014).

When comparing ESTs with genomic data especially those derived from whole-genome sequencing, usually ESTs are preferentially used to develop microsatellite markers (see review by Ellis & Burke 2007). EST-microsatellites have many advantages over genomic microsatellites,

such as a high level of cross-species portability and cleaner results for scoring owing to fewer null alleles (see review by Ellis & Burke 2007). For example, EST-microsatellites were found to be more than three times as transferable across species as compared with genomic microsatellites in *Helianthus* spp. (Pashley *et al.* 2006). These advantages of EST-microsatellites largely come from the conserved nature of flanking regions of EST-microsatellites (see review by Li *et al.* 2002 and references therein). Consistent with many studies, the EST-microsatellite markers developed in this study showed a low level of null alleles (Zhan *et al.* 2010, 2012). Under the circumstance of a highly polymorphic genome of *C. intestinalis*, such a low level of null alleles is derived from both careful selection and assessment of markers (such as removal of loci with poor amplification) and from the strategy of microsatellite development, that is microsatellite development through ESTs, rather than genomic resources (see below the section for more detail).

For *C. intestinalis*, sequence data from both whole-genome sequencing and EST projects are available (<http://genome.jgi.doe.gov/Cioin2/Cioin2.home.html>). Compared to microsatellites derived from ESTs, microsatellites mined from the whole-genome sequencing project were more abundant and relatively easily located on chromosomes. However, our trials showed that more than 68.5% of genomic microsatellites had poor amplification success across multiple individuals (48 of 70 loci randomly selected and tested, unpublished data). Such poor amplification was improved in EST-microsatellites, accounting for 24.4% of total characterized microsatellites (i.e. 75 of 308 loci, these loci were removed from our microsatellite panel, see the Results section). The poor amplification problem is likely due to a high level of genomewide polymorphism in *C. intestinalis*: as high as 1.2% based on single nucleotide polymorphisms and insertions/deletions as estimated by the whole-genome sequencing project (Dehal *et al.* 2002). Such a high level of polymorphism leads to frequent occurrence of mismatches in primer annealing sites (i.e. presence of null alleles), resulting in poor amplification success across individuals.

Poor cross-species utility within *C. intestinalis* species complex

Microsatellite markers have been well accepted to be transferable across related genera, especially for those with conserved flanking regions such as EST-microsatellites. For example, a test based on 34 loci showed that >99% loci could be successfully transferred among 24 bird species, including the pair between chicken and zebra finch, close to maximum divergence within birds (Dawson *et al.* 2010). However, an unpredictably low level of cross-amplification success was reported in some taxonomic groups, such as a 21% within genus amplifica-

tion success rate among ranid frog species (Primmer & Merilä 2002). Similarly, we detected a low cross-species utility rate: only 25.8% of microsatellite loci were transferable within *C. intestinalis* species complex in this study.

Although many factors can contribute to the success of cross-species utility of microsatellite markers (see review by Barbará *et al.* 2007), recent studies showed that there was a correlation between cross-species amplification success and genetic divergence between source-target species in several taxonomic groups, such as birds (Moodley *et al.* 2015), fish (Carreras-Carbonell *et al.* 2008), bats (Jan *et al.* 2012) and amphibians (Hendrix *et al.* 2010). In general, a markedly decreased amplification success was observed with increased phylogenetic distance from the source species. Primmer *et al.* (2005) conducted a comprehensive survey for birds, cetaceans and frogs, and found a similar trend across taxa, showing approximately 50% of the loci amplifying at a source-target species with divergence time of 8.5–24.0 million years ago (Ma). However, recent studies showed that the cross-species amplification success rate varied among taxonomic groups. For example, a survey in bats showed very little success above 20 Ma divergence time (Jan *et al.* 2012); however, cross-amplification success rate of 65% was detected in salamander species with more than 30 Ma (Hendrix *et al.* 2010). For *C. intestinalis* species complex, phylogenetic studies indicated that the species complex might have radiated sometime during the late Miocene in the brackish Sarmatian (10.5–8.2 Ma) or Pontian (8.2–6.3 Ma) seas (Zhan *et al.* 2010). Based on the divergence time and estimation derived from related studies, we expected a cross-amplification success rate of >50%. However, the success rate of 25.8% obtained in this study was far below what we expected.

Several more factors may be responsible for such a low level of transferability of microsatellites among congeneric species for marine invertebrates. The unique biological characteristics, such as large population size and high mutation rate, could likely contribute to the low transferability. Despite that the correlation between cross-species amplification success and genetic divergence between source-target species was detected in several species such as oysters, the success rate decreased more sharply as the genetic divergence increased. For example, among 86 microsatellite markers originally developed for the Pacific oyster *Crassostrea gigas*, 83 (96.5%) were likely to be useful for the Portuguese oyster *C. angulata*, 71 (82.6%) for the Kumamoto oyster *C. sikamea*, 31 (36.0%) for the Suminoe oyster *C. ariakensis*, and only 11 (12.8%) for the Eastern oyster *C. virginica*. *C. angulata* is the closest relative of the

C. gigas with the divergence of several hundred thousand years, followed by *C. sikamea* (2.3 Ma) and *C. ariakensis* (3.6 Ma). For the pair of *C. gigas* and *C. virginica*, the divergence time was estimated as approximately 5 Ma, the success rate was as low as 12.8% (Hedgecock *et al.* 2004). Similarly, the low cross-utility rate at the genome level in *C. intestinalis* might be derived from high mutation rate in the AT-rich genome (~65% AT content; Dehal *et al.* 2002). In addition, *C. intestinalis* has a short generation time – usually two or three generations each year (Dybern 1965). The short generation time can potentially build up a high level of genetic divergence among congeneric species in a relatively short period of evolutionary time. Owing to these factors, mutations in flanking regions of microsatellites (i.e. mismatches in PCR primer annealing sites) can be common among related species, leading to a low level of cross-species amplification success. Consequently, both evolutionary and biological characteristics should be considered when estimating and testing cross-species utility of microsatellite markers among related species.

Conclusions

In this study, we successfully developed and characterized a large set of genomewide gene-associated microsatellite markers for a model invasive ascidian, *Ciona intestinalis*. These genetic markers were relatively evenly distributed across the whole genomes. As *C. intestinalis* has become a good model for the study of microevolution associated with rapidly changing environments in the wild, the large number of loci developed here, as well as the large proportion of the genome represented, makes this marker set a valuable resource for genomewide survey to identify genomic regions and/or genes under selection.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31272665) and 100-Talent Program of the Chinese Academy of Sciences to AZ.

References

- Alía R, Chambel R, Notivol E, Climent J, González-Martínez SC (2014) Environment-dependent microevolution in a Mediterranean pine (*Pinus pinaster* Aiton). *BMC Evolutionary Biology*, **14**, 200.
- Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, **16**, 3759–3767.
- Björklund M, Ranta E, Kaitala V, Bach LA, Lundberg P, Stenseth NC (2009) Quantitative trait evolution and environmental change. *PLoS ONE*, **4**, e4521.

- Brunetti R, Gissi C, Pennati R, Caicci F, Gasparini F, Manni L (2015) Morphological evidence that the molecularly determined *Ciona intestinalis* type A and type B are different species: *Ciona robusta* and *Ciona intestinalis*. *Journal of Zoological Systematics and Evolutionary Research*, **53**, 186–193.
- Caputi L, Andreakis N, Mastrototaro F, Cirino P, Vassillo M, Sordino P (2007) Cryptic speciation in a model invertebrate chordate. *Proceedings of the National Academy of Sciences*, **104**, 9364–9369.
- Carlton JT, Ruiz GM (2005) Vector science and integrated vector management in bioinvasion ecology: conceptual frameworks. In: *Invasive Alien Species: A new Synthesis* (eds Mooney HA, Mack RN, McNeely JA, Neville LE, Schei PJ, Waage JK), pp. 36–58. Covelo, Island Press, California.
- Carreras-Carbonell J, Macpherson E, Pascual M (2008) Utility of pairwise mtDNA genetic distances for predicting cross-species microsatellite amplification and polymorphism success in fishes. *Conservation Genetics*, **9**, 181–190.
- Carver CE, Chisholm A, Mallet AL (2003) Strategies to mitigate the impact of *Ciona intestinalis* (L.) biofouling on shellfish production. *Journal of Shellfish Research*, **22**, 621–631.
- Charmantier A, Garant D (2005) Environmental quality and evolutionary potential: lessons from wild populations. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 1415–1425.
- Collins S, de Meaux J (2009) Adaptation to different rates of environmental change in *Chlamydomonas*. *Evolution*, **63**, 2952–2965.
- Dawson DA, Horsburgh GJ, Küpper C *et al.* (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility – as demonstrated for birds. *Molecular Ecology Resources*, **10**, 475–494.
- Dehal P, Satou Y, Campbell RK *et al.* (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*, **298**, 2157–2167.
- Dybern BI (1965) The life cycle of *Ciona intestinalis* (L.) f. typica in relation to environmental temperature. *Oikos*, **16**, 109–131.
- Dybern BI (1967) Settlement of sessile animals on eternite slabs in two polls near Bergen. *Sarsia*, **29**, 137–180.
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity*, **99**, 125–132.
- Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167–178.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices. Version 2.9.3.2. Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*, **48**, 361–372.
- Hansen MM, Olivieri I, Waller DM, Nielsen EE, GeM Working Group (2012) Monitoring adaptive genetic responses to environmental change. *Molecular Ecology*, **21**, 1311–1329.
- Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research*, **23**, 379–385.
- Hendrix R, Hauswaldt JS, Veith M, Steinfartz S (2010) Strong correlation between cross-amplification success and genetic distance across all members of 'True Salamanders' (Amphibia: Salamandridae) revealed by *Salamandra salamandra*-specific microsatellite loci. *Molecular Ecology Resources*, **10**, 1038–1047.
- Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, **46**, 10–18.
- Jan C, Dawson DA, Altringham JD, Burke T, Butlin RK (2012) Development of conserved microsatellite markers of high cross-species utility in bat species (Vespertilionidae, Chiroptera, Mammalia). *Molecular Ecology Resources*, **12**, 532–548.
- Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, **23**, 1683–1685.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, **11**, 2453–2465.
- Lin Y, Gao Z, Zhan A (2015) Introduction and use of non-native species for aquaculture in China: status, risks and management solutions. *Reviews in Aquaculture*, **75**, 28–58.
- Moodley Y, Masello JF, Cole TL *et al.* (2015) Evolutionary factors affecting the cross-species utility of newly developed microsatellite markers in seabirds. *Molecular Ecology Resources*, **15**, 1046–1058.
- Navajas-Pérez R, Robles F, Molina-Luzón MJ *et al.* (2012) Exploitation of a turbot (*Scophthalmus maximus* L.) immune-related expressed sequence tag (EST) database for microsatellite screening and validation. *Molecular Ecology Resources*, **12**, 706–716.
- Nydam ML, Harrison RG (2007) Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). *Marine Biology*, **151**, 1839–1847.
- Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. *Journal of Heredity*, **97**, 381–388.
- Primmer CR, Merilä J (2002) A low rate of cross-species microsatellite amplification success in Ranid frogs. *Conservation Genetics*, **3**, 445–449.
- Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology*, **36**, 348–360.
- Procaccini G, Affinito O, Toscano F, Sordino P (2011) A New Animal Model for Merging Ecology and Evolution. In: *Evolutionary Biology-Concepts, Biodiversity, Macroevolution and Genome Evolution* (ed. Pontarotti P), pp. 91–106. Springer-Verlag, Berlin.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Richardson DM, Rejmánek M (2011) Trees and shrubs as invasive alien species—a global review. *Diversity and Distributions*, **17**, 788–809.
- Satoh N, Satou Y, Davidson B, Levine M (2003) *Ciona intestinalis*: an emerging model for whole-genome analyses. *Trends in Genetics*, **19**, 376–381.
- Satou Y, Mineta K, Ogasawara M *et al.* (2008) Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biology*, **9**, R152.
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.
- Shi J, Huang S, Zhan J *et al.* (2014) Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. *DNA Research*, **21**, 53–68.
- Therriault TW, Herborg LM (2008) Predicting the potential distribution of the vase tunicate *Ciona intestinalis* in Canadian waters: informing a risk assessment. *ICES Journal of Marine Science: Journal du Conseil*, **65**, 788–794.
- Valentine PC, Carman MR, Blackwood DS, Heffron EJ (2007) Ecological observations on the colonial ascidian *Didemnum* sp. in a new England tide pool habitat. *Journal of Experimental Marine Biology and Ecology*, **342**, 109–121.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, **93**, 77–78.
- Waters JM, Dijkstra LH, Wallis GP (2000) Biogeography of a southern hemisphere freshwater fish: how important is marine dispersal? *Molecular Ecology*, **9**, 1815–1821.
- Zhan A, Macisaac HJ, Cristescu ME (2010) Invasion genetics of the *Ciona intestinalis* species complex: from regional endemism to global homogeneity. *Molecular Ecology*, **19**, 4678–4694.

- Zhan A, Darling JA, Bock DG, Lacoursière-Roussel A, MacIsaac HJ, Cristescu ME (2012) Complex genetic patterns in closely related colonizing invasive species. *Ecology and Evolution*, **2**, 1331–1346.
- Zhan A, Briski E, Bock DG, Ghabooli S, MacIsaac HJ (2015) Ascidiaceans as models for studying invasion success. *Marine Biology*. doi:10.1007/s00227-015-2734-5.

A.Z. conceived the project, Y.L., Y.C. and W.X. conducted the laboratory work, Y.L. analyzed the data, Y.L. and A.Z. wrote the manuscript, and all authors reviewed and commented on the manuscript.

Data accessibility

All EST-microsatellites were mined from the public database at the Joint Genome Institute (JGI) website at <http://genome.jgi.doe.gov/Cioin2/Cioin2.home.html>. The chromosomal locations of the 2126 gene-associated microsatellites for *Ciona intestinalis* spA have been uploaded as online supplemental material (Appendix S3). All raw genotyping data for both *Ciona*

intestinalis spA and spB were deposited at DRYAD entry doi: 10.5061/dryad.dm98k.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Characterization of 218 microsatellite loci for *Ciona intestinalis* spA.

Appendix S2 Characterization of 60 successfully cross-amplified microsatellite loci for *Ciona intestinalis* spB.

Appendix S3 The chromosomal location of gene-associated microsatellites for *Ciona intestinalis* spA.

Appendix S4 The chromosomal locations of the 163 gene-associated polymorphic microsatellite markers for *Ciona intestinalis* spA.

Appendix S5 The locations of the 41 gene-associated polymorphic microsatellite markers for *Ciona intestinalis* spA on each scaffold based on the KH assembly of Satou *et al.* (2008) available from the Ensembl website (http://asia.ensembl.org/Ciona_intestinalis/Info/Index?db=core).