Conservation Genet Resour (2014) 6:289–291
DOI 10.1007/s12686-013-0127-y

TECHNICAL NOTE

Development of anonymous nuclear markers from Illumina paired-end data for Seychelles caecilian amphibians (Gymnophiona: Indotyphlidae)

Claire J. Lewis · Simon T. Maddock · Julia J. Day · Ronald A. Nussbaum · Charles Morel · Mark Wilkinson · Peter G. Foster · David J. Gower

Received: 3 December 2013 / Accepted: 25 December 2013 / Published online: 11 January 2014

Abstract Anonymous nuclear markers were developed for Seychelles caecilian amphibians. Using a previously published bioinformatics pipeline (developed for Roche 454 data), 36 candidate anonymous nuclear loci (ANL) of at least 180 bp length were identified from Illumina MiSeq next generation sequencing data for five Seychelles species. We designed primer pairs for the 36 candidate ANL and tested these by PCR and Sanger sequencing. Seven ANL amplified and sequenced well for at least five of the six nominal Seychelles caecilian species (in three genera), and represent potentially useful markers for systematics and conservation.

Keywords Anonymous nuclear loci · Conservation · Next generation sequencing · Population genetics · Systematics

Electronic supplementary material The online version of this article (doi:10.1007/s12686-013-0127-y) contains supplementary material, which is available to authorized users.

C. J. Lewis · S. T. Maddock (✉) · M. Wilkinson · P. G. Foster · D. J. Gower
Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK
e-mail: s.t.maddock@gmail.com

C. J. Lewis · S. T. Maddock · J. J. Day · M. Wilkinson
Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

R. A. Nussbaum
Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA

C. Morel
Natural History Museum, Victoria, Mahé, Seychelles

Anonymous nuclear loci (ANL) are markers in non-coding regions of the nuclear genome that are unlikely to be under selection and which have many potential applications in molecular systematics, especially at lower taxonomic levels (Karl and Avise 1993). Traditionally, ANL discovery involved time-consuming lab-work to generate a small number of usable markers, but recently Bertozzi et al. (2012) described a bioinformatics pipeline to develop candidate ANL from next generation sequencing (NGS) data.

Caecilians (Gymnophiona) are limbless, mostly soil-dwelling amphibians largely restricted to the moist tropics, and the approximately 200 extant species comprise one of the most poorly known major vertebrate groups (Gower and Wilkinson 2008; Wilkinson 2012). Anonymous nuclear markers have not previously been developed for caecilians, microsatellites have been developed for only two species (Li et al. 2010; Barratt et al. 2012), and published coding nuclear data are not very variable at lower taxonomic levels.

A radiation of six nominal caecilian species in three genera (*Grandisonia, Hypogeophis*, *Praslinia*) occurs in the Seychelles (Nussbaum 1984; Wilkinson and Nussbaum 2006; Wilkinson et al. 2011). Although clearly monophyletic (Nussbaum and Ducey 1988; Hedges et al. 1993; Gower et al. 2011), analyses of (mostly mtDNA) sequence data have been unable to robustly resolve all relationships among the radiation (Hedges et al. 1993; Wilkinson et al. 2002, 2003; Loader et al. 2007; Gower et al. 2008, 2011). ANL could provide a useful tool for Seychelles caecilian systematics and conservation genetics, especially given the increased levels of threat faced, in general, by island biotas (e.g., Frankham 2008). Two Seychelles caecilians (*P. cooperi* and *H. brevis*) are classified as Endangered on the IUCN Red List.

Six samples from five Seychelles caecilian species (all Seychelles species except *P. cooperi*; two samples of *G. alternans*) were selected for marker development using NGS. Genomic DNA was extracted from liver using Qiagen DNeasy Blood and Tissue kits. Samples were prepared using a standard Illumina Nextera DNA kit and paired-end reads (≤251 bp long) sequenced using a 500 cycle v.2 reagent kit on the Illumina MiSeq platform. Paired-end data were combined and cleaned using default settings in Geneious v.6.1.4. Cleaned files were run through Bertozzi et al.'s (2012) Perl bioinformatics pipeline, with our customised BLAST database comprising the *Xenopus tropicalis* and *Danio rerio* genomes plus all caecilian entries in GenBank. Each sample was run through the pipeline individually (see supplementary material). Candidate anonymous nuclear markers were selected at random from paired reads ≥245 bp, and 36 primer pairs (6 per sample) were designed using Primer3 v.0.4.0 (Koressaar and Remm 2007; Untergrasser et al. 2012). Primer sequences were selected on the basis of being closest to the beginning of the 5′–3′ end of each read (to maximise length), and having low self-complementarity and annealing temperatures of 60 °C (±1 °C); all other settings were default. Primer sequences were subjected to an additional BLAST search to check locus anonymity.

The 36 primer pairs were tested using the polymerase chain reaction (PCR) for five genomic DNA samples, one for each of the non-*P. cooperi* Seychelles species (see supplementary material). Reaction volume was 25 µl: 1 µl template, 1 µl for each primer, 9.5 µl of dd H_2O, 12.5 µl of MyTaq Mix ×2. Cycling conditions for all primer pairs were: 95 °C-3 min; 35× [95 °C-15 s, 60 °C-15 s, 72 °C-20 s]; 72 °C-10 min. Fifteen of the 36 primer pairs successfully amplified DNA in all five species, and amplicons for these were subjected to Sanger sequencing. Assembled and edited sequences were aligned using default settings for consensus alignments in Geneious. Eight of the 15 loci were considered suitable for further testing; the seven loci rejected at this stage generally yielded poor sequences, perhaps indicative of suboptimal primer/template combinations and/or PCR settings. Sequences were subjected to a BLAST search to check anonymity.

The eight surviving candidate ANL were tested further by attempting PCR amplification of genomic DNA from 12 additional individuals of the five species for which they had already worked plus two individuals of their Seychelles sister species *P. cooperi* (see supplementary material). Descriptive statistics were generated using DnaSP v.5.10 (Librado and Rozas 2009) for seven ANL (see Table 1); all were variable across the Seychelles species. One locus was excluded from DnaSP analysis because it failed to sequence well in any specimen of *H. rostratus* or *G. larvata*.

Our NGS data initially produced approximately 5,000 candidate ANL with a potential length of ≥245 bp. Approximately 20 % of the 36 candidate ANL that we subsequently randomly selected were found to generally amplify and sequence well and showed variability in the

Table 1 The seven anonymous nuclear markers developed successfully in this study

Locus	Primer sequences 5′–3′	Length (range)	SA (SS)	Spp.	PS	Indels (range)	VS %
Alt15	F: GCCTTGCATCCCCTAATACA	283 (137–283)	17 (11)	6	24	6 (1–24)	8.5
	R: GCACACACTGTCGGCTTAAA						
Alt23	F: TCCATAGGAAGGGAGCAAGA	299 (277–298)	16 (10)	5	24	4 (1–14)	8.4
	R: CTGCCCGCTTTCTTTGTAAC						
Brev2	F: TAGAAGCCGAGGGTTATTGG	199 (187–199)	19 (9)	5	21	1 (8)	11.1
	R: GAAGAGAAGGTGGGACAGGA						
Brev5	F: CATCAGGTCATTGGCGTTTA	324 (273–323)	17 (10)	6	35	3 (1)	11.1
	R: GAGTGCAGGGACCAAATACC						
Rost1	F: TCTGGAATTGGCCTTGTGTT	291 (288–291)	15 (7)	5	22	3 (1–2)	8.6
	R: CCCACATTCTTCCTCCCTCT						
Rost5	F: TGTCAACTGCCCTCTGTGTC	326 (319–325)	14 (11)	6	24	5 (1–2)	8.3
	R: AAATTCACAGGCCAAACAGG						
Sech5	F: GCAGCTCTTTCTGTGCCTTT	180	16 (10)	6	25	0	13.9
	R: GTCTGCCATTGCTGTATGGA						

For each locus we report (columns, from left to right) locus name; primers; length in base pairs of aligned sequences; the number of genomic DNA samples out of 19 for which PCR amplification worked (SA) and for which sequence data were successfully obtained (SS); the number of species out of six for which amplificiations were achieved and for which amplicons were sequenced successfully (Spp.); the number of polymorphic sites (PS); the number of indels across the alignment (and per sequence, in parentheses); the percentage of variable sites excluding indels (VS %). See online resources for additional details

Seychelles caecilians. Our successful approach differed from Bertozzi et al.'s (2012) in that we developed ANL from Illumina NGS data rather than from the more expensive Roche 454 platform, albeit at the expense of sequence length.

Acknowledgments For help with office, lab and fieldwork we thank Rachel Bristol, Pat Dyal, Jim Labisko, Greg Schneider, Bruno Simões, Martijn Timmermans and Gill Sparrow. The Seychelles Bureau of Standards provided field collection and export permits. STM thanks the Seychelles National Park Authority for supporting his application for a gainful occupancy permit. Funded in part by an NHM/UCL PhD studentship and Systematics Association Research Fund Grant (to STM), and a SynTax grant and Darwin Initiative grant 19-002 (to MW, JJD and DJG).

References

Barratt CD, Horsburgh GJ, Dawson DA, Gower DJ, Wilkinson M, Loader SP, Jehle R (2012) Characterisation of nine microsatellite loci in the caecilian amphibian *Boulengerula uluguruensis* (Gymnophiona), and their cross-species utility in three congeneric species. Conserv Genet Resour 4:225–229

Bertozzi T, Sanders KL, Sistrom MJ, Gardner MG (2012) Anonymous nuclear loci in non-model organisms: making the most of high throughput genome surveys. Bioinformatics 28:1807–1810

Frankham R (2008) Inbreeding and extinction: island populations. Conserv Biol 12:665–675

Gower DJ, Wilkinson M (2008) Caecilians (Gymnophiona). In: Stuart SN, Hoffman M, Chanson JS, Cox NA, Berridge R, Ramani P, Young BE (eds) Threatened amphibians of the world. Lynx Edicones with IUCN—The World Conservation Union. Conservation International and NatureServe, Barcelona, pp 19–20

Gower DJ, Giri V, Dharne MS, Shouche Y (2008) Frequency of independent origins of viviparity among caecilians (Gymnophiona): evidence from the first 'live-bearing' Asian amphibian. J Evol Biol 2:1220–1226

Gower DJ, San Mauro D, Giri V, Bhatta G, Venu G, Ramachandran K, Oommen OV, Fatih FA, Mackenzie-Dodds JA, Nussbaum RA, Biju SD, Shouche TS, Wilkinson M (2011) Molecular systematics of caeciliid caecilians (Amphibia: Gymnophiona) of the Western Ghats, India. Mol Phylogenet Evol 59:698–707

Hedges SB, Nussbaum RA, Maxson LR (1993) Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12S rRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol Monogr 7:64–76

Karl SA, Avise JC (1993) PCR-based assays of mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. Mol Biol Evol 10:342–361

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291

Li G, Luo X, Wang H, Bei Y, Shi W, Meng S, Zhang B (2010) Isolation and characterization of fifteen polymorphic tetranucleotide microsatellites for *Ichthyophis bannanicus* (Gymnophiona: Ichthyophiidae). Conserv Genet Resour 2:129–131

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Loader SP, Pisani D, Cotton JA, Gower DJ, Day JJ, Wilkinson M (2007) Relative time scales reveal multiple origins of parallel disjunct distributions of African caecilian amphibians. Biol Lett 3:505–508

Nussbaum RA (1984) Amphibians of the Seychelles. In: Stoddart DR (ed) Biogeography and ecology of the Seychelles islands. Dr. W. Junk, The Hague, pp 379–415

Nussbaum RA, Ducey PK (1988) Cytological evidence for monophyly of the caecilians (Amphibia: Gymnophiona) of the Seychelles archipelago. Herpetologica 44:290–296

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115

Wilkinson M (2012) Caecilians. Curr Biol 22:R668–R669

Wilkinson M, Nussbaum RA (2006) Caecilian phylogeny and classification. In: Exbrayat J-M (ed) Reproductive biology and phylogeny of Gymnophiona (caecilians). Science Publishers Inc., Enfield, New Hampshire, pp 39–78

Wilkinson M, Sheps JA, Oommen OV, Cohen BL (2002) Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol Phylogenet Evol 23:401–407

Wilkinson M, Loader SP, Gower DJ, Sheps JA, Cohen BL (2003) Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr J Herpetol 52:83–92

Wilkinson M, San Mauro D, Sherratt E, Gower DJ (2011) A nine-family classification of caecilians (Amphibia: Gymnophiona). Zootaxa 2874:41–64