

Design and Implementation of Degenerate Microsatellite Primers for the Mammalian Clade

Emmanuel Buschiazzo^{*†a}, Josephine S. Beck, Neil J. Gemmell^{†b}

School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

Abstract

Microsatellites are popular genetic markers in molecular ecology, genetic mapping and forensics. Unfortunately, despite recent advances, the isolation of *de novo* polymorphic microsatellite loci often requires expensive and intensive groundwork. Primers developed for a focal species are commonly tested in a related, non-focal species of interest for the amplification of orthologous polymorphic loci; when successful, this approach significantly reduces cost and time of microsatellite development. However, transferability of polymorphic microsatellite loci decreases rapidly with increasing evolutionary distance, and this approach has shown its limits. Whole genome sequences represent an under-exploited resource to develop cross-species primers for microsatellites. Here we describe a three-step method that combines a novel *in silico* pipeline that we use to (1) identify conserved microsatellite loci from a multiple genome alignments, (2) design degenerate primer pairs, with (3) a simple PCR protocol used to implement these primers across species. Using this approach we developed a set of primers for the mammalian clade. We found 126,306 human microsatellites conserved in mammalian aligned sequences, and isolated 5,596 loci using criteria based on wide conservation. From a random subset of ~1000 dinucleotide repeats, we designed degenerate primer pairs for 19 loci, of which five produced polymorphic fragments in up to 18 mammalian species, including the distinctly related marsupials and monotremes, groups that diverged from other mammals 120–160 million years ago. Using our method, many more cross-clade microsatellite loci can be harvested from the currently available genomic data, and this ability is set to improve exponentially as further genomes are sequenced.

Citation: Buschiazzo E, Beck JS, Gemmell NJ (2011) Design and Implementation of Degenerate Microsatellite Primers for the Mammalian Clade. *PLoS ONE* 6(12): e29582. doi:10.1371/journal.pone.0029582

Editor: Robert C. Fleischer, Smithsonian Institution National Zoological Park, United States of America

Received: May 25, 2011; **Accepted:** December 1, 2011; **Published:** December 27, 2011

Copyright: © 2011 Buschiazzo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Royal Society of New Zealand Marsden grant (UOC 202) to NJG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: elbuzzo@gmail.com

†a Current address: School of Natural Sciences, University of California Merced, Merced, California, United States of America

†b Current address: Department of Anatomy and Structural Biology, Centre for Reproduction and Genomics, University of Otago, Dunedin, New Zealand

Introduction

Microsatellites, also called simple sequence repeats, consist of short (1–6 bp), tandemly repeated DNA motifs dispersed throughout genomes. Microsatellite sequences mutate through motif insertions and deletions along the repeat array, often at rates several orders of magnitude higher than the average genomic mutation rate [1]. Increasing numbers of polymorphic microsatellites are being associated with genetic disorders and variation in gene expression [2], but the high mutation rate at microsatellite loci also offers an abundant and readily available polymorphism that has been the foundation for the wide use of microsatellites as neutral molecular markers, especially in applications requiring fine temporal and/or spatial resolution, e.g. population genetics and forensics.

Despite a number of recognized advantages of microsatellites over other genetic markers, such as easy sample preparation and high information content [3,4], the costs and time required to develop new polymorphic microsatellite markers can be prohibitive [5]. The recent decline in sequencing costs has paved the way for more efficient methods of *de novo* microsatellite isolation, but only when whole genome sequences [6] or large amounts of sequences are already available [7,8] or purposely produced for

the species of interest [9]; conditions that still imply a significant upstream investment.

Seeking to yield large amounts of genetic information with the least initial effort and cost, investigators commonly make attempts at transferring known microsatellite markers between species, typically from previously examined focal species to related non-focal species (e.g. [10,11]; see [12] for a recent review). Successful transfer of microsatellite markers therefore requires (i) 1:1 orthology of microsatellite loci, (ii) flanking sequences which are sufficiently conserved between species to provide PCR priming sites for cross-species amplification, and (iii) a microsatellite sequence which exhibits an appropriate level of polymorphism in the non-focal species. All three aspects are typically unknown at the onset of a project. Because there is a strong positive relationship between time of divergence and the accumulation of sequence differences along lineages [13], the consensus found in the literature that microsatellite transferability rapidly decreases with increasing distance between focal and non-focal species is not surprising [11,12,14–16].

With no prior focus on reducing the impact of these limitations, the traditional cross-species microsatellite transfer approach has had varying, generally disappointing, levels of success [12]. The use of databases of microsatellites located in expressed, thus

putatively conserved genomic regions, have improved the expected and observed rate of cross-species transferability (e.g. [17–21]), especially with the complementary use of genome sequences from related species [22,23]. However, multiple whole-genome alignments have not yet been exploited to date to explore and maximize the limits of microsatellite marker transferability.

Here, we present a novel and economic strategy that exploits our recent advances in building comprehensive datasets of microsatellites conserved across the mammalian clade [24–26]. We created a reproducible and adaptable framework that has allowed us to develop mammal-wide degenerate primers for nine dinucleotide microsatellites, five of which were successfully genotyped across most of a panel comprising 18 divergent species that represent the major mammalian orders, and three of which displayed high intraspecies polymorphism throughout the mammals tested. We conclude from this successful initial trial that this

approach has much promise and paves the way for equivalent studies in other genera as the push towards obtaining genome sequences from multiple animal, predominantly vertebrate, species becomes a reality [27]. In addition, it provides a significant starting resource for those wishing to focus on specific mammalian species or groups of species where large numbers of microsatellite markers with robust cross-species utility are required.

Materials and Methods

Our overall strategy is presented in Figure 1.

Ethics Statement

Information S1 shows the origin of our samples for each species included in this study. Restricted and general biological products (tissue or DNA) were imported with the New Zealand Ministry of

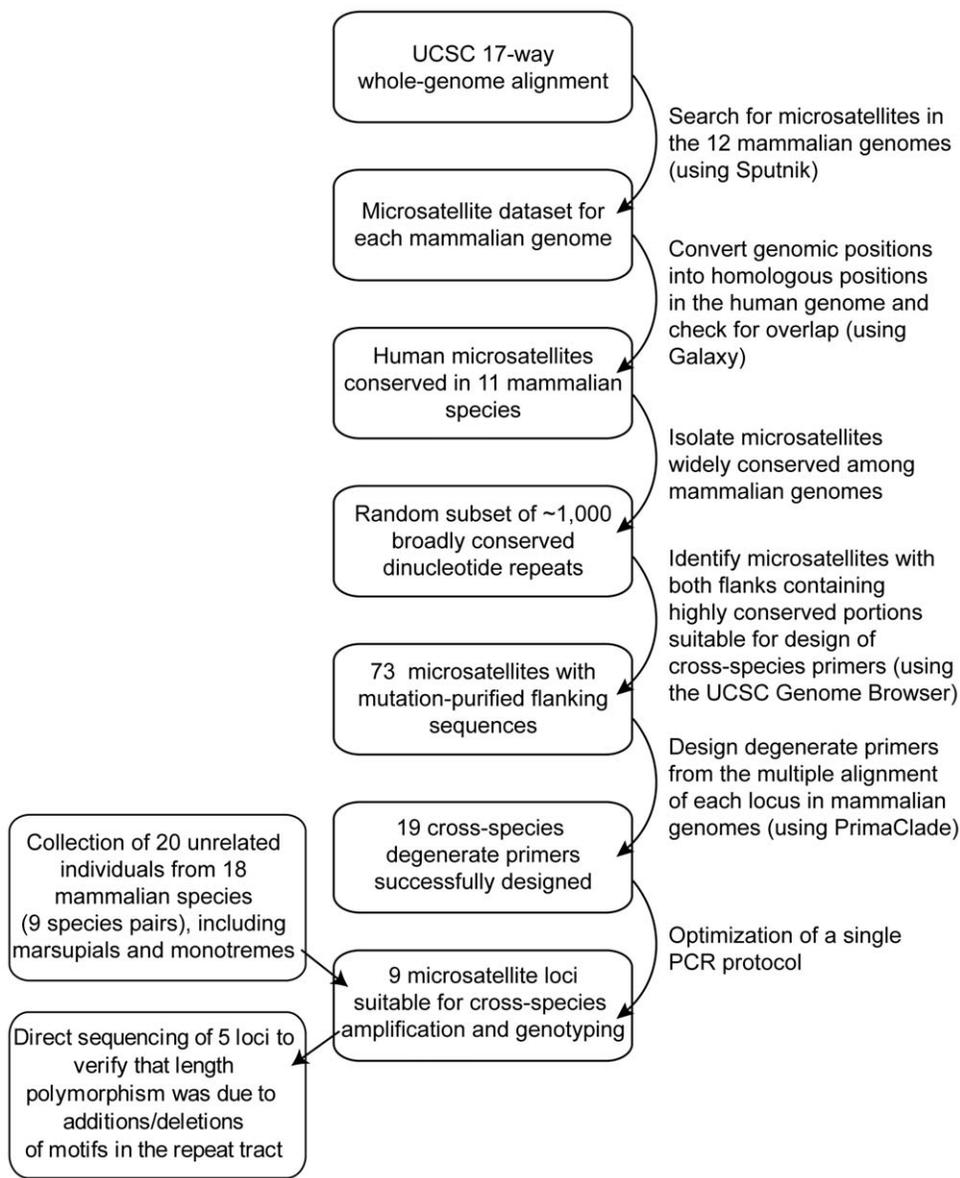


Figure 1. Schematic representation of the pipeline developed to design and implement degenerate cross-species primers for mammal-wide microsatellite loci. The University of California, Santa Cruz (UCSC) Genome Browser can be found at <http://genome.ucsc.edu/>. doi:10.1371/journal.pone.0029582.g001

Agriculture and Forestry (MAF) import permits No 2007031396 and 2007032360, respectively, issued for the University of Canterbury. Human DNA was sent from the National Cell Bank of Iran. Chimpanzee samples Pt163, Pt180, Pt203, and Pt254 were obtained as blood samples from the Iberia Research Center during routine veterinary care, and were processed in A. Stone's laboratory in Arizona State University, USA.

Sample collection and DNA extraction

Mammalian species were chosen to include nine sister species pairs ($n = 18$) representing three of the four superorders of eutherians (Laurasatheria, Euarchontoglires and Afrotheria), as well as the too often neglected marsupials and monotremes. We collected DNA, blood or tissue samples from 20 presumably unrelated individuals per species (Information S1). Total DNA was extracted using slight variations of the Chelex method [28], and quantified using an ND-1000 spectrophotometer (NanoDrop).

In silico identification of conserved mammalian microsatellites

Orthologous mammalian microsatellites were identified using the UCSC vertebrate 17-WA [29] and a variation of an approach detailed elsewhere [26]. Briefly, FASTA-formatted sequences were extracted from the alignment in a pairwise fashion (human-other species) using Gmaj (<http://globin.cse.psu.edu/dist/gmaj/>), and the microsatellite search was carried out with a modified version of Sputnik [30], using the following parameters: $-v 1 -u 5 -n -4 -s 8 -L 15$ (motif length: 1–5 bp; mismatch penalty: -4; min score: 8, min array length: 15 bp). Each dataset was filtered for single-copy and repeat-free loci, and classified according to motif type, length, purity and complexity. Genomic positions of non-human microsatellites were converted to homologous positions in the human genome using a stand-alone version of Galaxy [31] and resources available at the UCSC Genome Browser (i.e. liftOver tool and conversion files). Conservation was assigned when genomic positions of human microsatellites overlapped with converted positions of non-human microsatellites.

In silico isolation of potential cross-species microsatellite loci

An initial subset of $\sim 1,000$ human dinucleotide microsatellites (length ≥ 14 bp) was randomly selected from a pool of broadly conserved microsatellites in the mammalian clade, i.e. present in at least in five mammals, or in comparisons including at least human, opossum and either dog or mouse. Mammalian species included in this study shared a common ancestor 160 MYA, and thus the chances of finding conserved *and* polymorphic were expected to be low. Although microsatellites composed of larger motifs, e.g. tri- and tetranucleotide repeats, are known to be less prone to genotyping errors than dinucleotide repeats, we chose to look and test the latter over the former because they tend to be longer and thus more polymorphic [32]. We did not require that microsatellites in this initial subset be conserved in all nine non-primate species for two reasons. First, most genome sequences in the 17-WA are incomplete, thus there is a non-negligible possibility for false negatives. Second, otherwise conserved microsatellites may be too short or overly interrupted to be detected using our *in silico* strategy in some genomes.

In order to optimize the identification of cross-species microsatellites with flanking sequences conserved across the entire mammalian clade, including monotremes (platypus), we reviewed by eye each microsatellite locus in the 28-way conservation track [33], an updated and enlarged version of the 17-WA (Information

S2). Criteria for selection were: (i) presence of a dinucleotide repeat in all taxa included in our sample collection; although exceptions were tolerated for low-coverage ($2\times$) genomes (cat, armadillo, elephant and tenrec), (ii) a relative extent of interspecies length variation in the repeat array, i.e. microsatellites with no or very limited length variation between species were discarded, (iii) ~ 20 or more near-identical contiguous base pairs on both sides of the microsatellite sequence across all mammalian species, and (iv) total length of the potential amplicon not exceeding ~ 400 bp. The purpose of this process, which by nature was relaxed because it was carried out by eye, was to filter out those microsatellites that did not meet the general requirements for cross-species markers, i.e. no variation in motif, but variation in length across species, conserved potential primer sites on both sides of the microsatellite and length of the amplicon compatible with current genotyping technology. We removed from these alignments any sequence derived from species not included in our sample collection, with the exception of sequences from armadillo (Xenarthra), elephant (Afrotheria) and opossum (Marsupialia), ensuring that each alignment covered the entire breadth of the Mammalia. When necessary, microsatellite flanking sequences were re-aligned manually using BioEdit [34].

In silico comparative primer design

Alignments were submitted to PrimaClade [35]; this web application runs Primer3 [36] independently for each sequence, collating the results to identify primers that bind across the alignment, while allowing for base degeneracy. A maximum of three degenerate sites per primer were allowed. Primers that overlapped gaps (indels) in the alignment were excluded, and only primers generating fragments smaller than 350 bp were kept for further study. Using the Java web-application NetPrimer and the developer's recommendations (PREMIER Biosoft International, <http://www.premierbiosoft.com/netprimer/>), potential primer pairs were tested for the presence of secondary structures (hairpins, self- and cross-dimerization), palindromes and repeats that could affect the amplification reaction through intra- and intermolecular interactions and non-specific annealing. Table 1 summarizes the overall set of unambiguous criteria that were applied to increase chances of successful amplification and select the optimal cross-species primer pair at each locus. In addition, the same criteria were used to design primers for a locus containing the non-coding microsatellite with the widest range of conservation in mammals described to date, and located in the 3'-UTR of the NCAM1 gene [37]. A list of all degenerate cross-species primers and their characteristics is displayed in Information S3.

DNA amplification, genotyping and sequencing

We followed the M13-tail PCR method of [38] and optimized it for cross-species investigation. Amplifications were performed on a Master Cycler Gradient (Eppendorf), in $15 \mu\text{l}$ of reactions containing $0.66 \mu\text{M}$ of reverse-specific primer, $0.66 \mu\text{M}$ of fluorescent dye-labelled M13 primer, $0.33 \mu\text{M}$ of forward-specific primer with M13-tail, 2.5 mM of MgCl_2 , 0.2 mM of each dNTP, 4 mM of tetramethylammonium chloride (TMAC), 0.75 U of BioTaq DNA polymerase (Bioline), and $20\text{--}100 \text{ ng}$ of genomic DNA template. A touch-down PCR was undertaken in which the initial annealing temperature T_{init} (generally 59°C , but see exceptions in Information S3) was reduced at the rate of 2°C every two PCR cycles until the target temperature ($T_{\text{targ}} = T_{\text{init}} - 10^\circ\text{C}$) was reached; 26 regular cycles were then performed at T_{targ} . The general thermocycling profile was as follows: initial denaturation at 94°C for 3 minutes; denaturation at 94°C for 15 s, annealing for 30 s, extension at 72°C for 20 s; final extension at 72°C for 20 minutes. PCR efficiency was assessed through electrophoresis of

Table 1. Selection criteria for designing comparative microsatellite primers.

L_{expected}	$L_{\text{primer}}^{\dagger}$	T_m^{\ddagger}	ΔT_m	%GC [§]	Repeats		Stability of primer secondary structures (ΔG^*)					
					2-6 ×	1 ×	3' HP	Int HP	3' SD	Int SD	3' CD	Int CD
<350	18–22	58–62	<1	45–60	<3	<6	>–2.00	>–3.00	>–5.00	>–6.00	>–5.00	>–6.00

L_{expected} : expected length of PCR products (bp); L_{primer} : primer length (bp); T_m : melting temperature ($^{\circ}\text{C}$); ΔT_m : T_m difference between both primers; %GC: G+C content; 2-6 ×: number of tandemly repeated non-monomer motifs (2–6 bp); 1 ×: length of mononucleotide runs; ΔG : Gibbs free energy required to break the secondary structure (kcal/mol); 3': 3'-end of primers; Int: Internal; HP: hairpin, SD: self-dimer, CD: cross-dimer.

*Output from NetPrimer; criteria as recommended in the application's manual.

[†]Exceptionally up to 26 bp.

[‡]Output from PrimaClade.

[§]A 30–62% range was tolerated for primers >22 bp.

doi:10.1371/journal.pone.0029582.t001

3 μl of amplified products loaded on 1.5% agarose gel stained with BET. Primer pairs resulting in multiple bands or no amplification in all or most species were discarded. Fragment analysis was performed in an ABI3100 Genetic Analyzer (Applied Biosystems) following instructions from the manufacturer. Fragment sizes were scored with GeneMarker (Soft Genetics LLC). Expected and observed heterozygosities and polymorphic information content (PIC) were measured for all genotyped loci in each species with CERVUS 3.0.3 [39,40]. We restricted sequencing to the loci where genotyping was successful in a broad range of species. Four individuals per species per locus were selected for direct sequencing on a locus per locus basis based on homozygosity and, where possible, polymorphism. The sequencing PCR was run using a standard protocol (Big Dye Terminator Cycle Sequencing Kit, Applied Biosystems), and products were prepared for sequencing in both directions in an ABI3100 Genetic Analyzer (Applied Biosystems) following manufacturer's instructions. Sequences obtained for each locus were aligned with ClustalW [41], and edited manually using BioEdit [34].

Results

Candidate cross-species microsatellite markers for the mammalian clade

A total of 126,306 human microsatellites were found conserved in at least one of the non-primate mammalian species, i.e. in mouse, rat, rabbit, dog, cow, elephant, armadillo, tenrec and/or opossum (Information S4). An initial subset of ~1,000 human dinucleotide microsatellites (length ≥ 14 bp) was randomly selected from a total pool of 5,596 microsatellites (including 2,756 dinucleotide repeats) that were broadly conserved across aligned genomes. Furthermore, a total of 73 28-WA intervals, each comprising a potential mammal-wide microsatellite locus, were selected for the presence of a polymorphic dinucleotide microsatellite flanked by stretches of ultra-conserved sequences potentially suitable for cross-species primer design (Information S2).

Degenerate primer pairs were then successfully designed for 19 microsatellite loci. Of those 19 primer pairs tested using a unique, optimized set of PCR conditions, nine pairs yielded a scorable band pattern in all tested mammalian samples (Information S5). There was no significant difference in amplification success between highly and slightly degenerate primer pairs, nor did primer G+C content of sequence affect amplification success (Information S3).

Intraspecies polymorphism

To test our set of mammal-wide microsatellite loci for length polymorphism at the population level, amplicons were produced

and genotyped in each 20-sample set. Of nine primer pairs developed for cross-species genotyping, five were successful in providing allele length data at the population level across most species. Table 2 shows allelic richness and estimates of heterozygosities (expected and observed) for each locus in each species, whenever genotyping was successful; Information S6 shows polymorphic information content (PIC). Although these values should be considered rough guides given the limited sample sizes, three microsatellite loci showed significantly more intraspecies polymorphism (C2-6868, C2-1915 and C2-1218), indicating potential suitability for marker-based applications across the mammalian clade.

It is interesting to weigh the extent of polymorphism at each locus against the sequence data that is available from the 28-way alignments. Indeed, intraspecies polymorphism is largely influenced by the length of pure repeat segments within the microsatellite sequence, with long pure microsatellite tracts tending to be more polymorphic than short and/or degenerated microsatellites [1]. Accordingly, the highly polymorphic C2-1218 locus contained long pure tracts of (CA) motifs in most species used for genotyping (Information S7). The C2-6868 locus showed less variability and contained many sub-units of short size (<8 repeats), with the exception of a long and extensively polymorphic tract in mouse (Information S7). Despite imperfections in the microsatellite sequence, the widely polymorphic C2-1915 locus generally contained at least one long pure sub-unit, i.e. >8 repeats (Information S7). The two other loci, C2-1514 and C17-4243, showed less polymorphism and generally contained short tracts (Information S7). Against expectations, we observed a few exceptionally long tracts with no intraspecies variability in allele length, for example the C17-4243 locus in rat. This may be explained by unintended close relatedness of individuals among some sample sets, e.g. rats and pilot whales (discussed below).

Relationship between changes in flanking sequences and locus length

Although sequencing is not standard practice in most applications of microsatellite markers, we sought to examine in detail the relationship between DNA sequence and the nature and extent of polymorphism of our most successful cross-species microsatellite loci across the studied species. Sequence-level information is indeed essential to inspect (i) whether allele length variations are attributable to additions/deletions of motifs within microsatellite sequences rather than indels in the flanking sequences, (ii) what is the extent of size homoplasy, if any, among alleles (homoplastic alleles have identical length but different sequence), (iii) the relationship between microsatellite structure and polymorphism [42–44]. Ideally, a microsatellite marker

Table 2. Polymorphism at nine cross-species mammalian microsatellite loci.

	C2-1218*	C2-6868*	C2-1915*	C4-1514*	C6-1112	C9-1918	C14-9692	C15-3531	C17-4243*
Human	268–294 (9/18) 0.47/0.82	228 (1/20) 0/0	166–178 (5/17) 0.71/0.64	281–283 (2/20) 0.30/0.26	152–156 (2/19) 0.11/0.19	300–302 (2/14) 0.38/0.52	234–237 (3/20) 0.05/0.15	226–228 (2/17) 0.29/0.26	311 (1/20) 0/0
Chimpanzee	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Mouse	291–301 (10/20) 0.30/0.76	242–291 (16/17) 0.56/0.93	216–238 (10/14) 0.50/0.91	313–317 (4/19) 0.22/0.52	155–161 (3/18) 0.50/0.51	311 (1/20) 0/0	240–242 (2/19) 0.26/0.56	297–299 (2/16) 0.13/0.12	319–325 (5/19) 0.42/0.62
Rat	274–280 (2/20) 0.45/0.36	236 (1/20) 0/0	176–180 (3/20) 0.35/0.31	274 (1/20) 0/0	158–162 (3/19) 0.47/0.55	n/a	n/a	237 (1/20) 0/0	326 (1/15) 0/0
Dog	268–278 (9/20) 0.40/0.83	256–268 (4/18) 0.28/0.52	180–191 (5/19) 0.47/0.78	297–299 (2/17) 0.12/0.11	3 peaks (1/11) n/a	n/a	214 (1/20) 0/0	n/a	309 (1/20) 0/0
Cat	265–276 (8/20) 0.75/0.82	n/a	176–188 (6/19) 0.63/0.72	n/a	n/a	n/a	n/a	n/a	312 (1/20) 0/0
Cow	259–264 (2/18) 0.06/0.06	231 (1/20) 0/0	167–169 (2/20) 0.15/0.22	281 (1/20) 0/0	146 (1/20) 0/0	300–305 (2/20) 0.50/0.51	208 (1/20) 0/0	240–242 (2/20) 0.05/0.05	308 (1/20) 0/0
Sheep	270–280 (8/19) 0.58/0.83	229–237 (4/14) 0.29/0.37	163–173 (4/15) 0.56/0.64	292 (1/20) 0/0	146 (1/20) 0/0	307–308 (2/20) 0/0.10	208–212 (3/18) 0.39/0.60	n/a	306 (1/20) 0/0
Dolphin	264–278 (4/19) 0.47/0.61	n/a	160–176 (7/19) 0.74/0.81	291–295 (2/16) 0.13/0.12	148–150 (2/19) 0.17/0.25	313–319 (3/16) 0.50/0.59	214–215 (2/19) 0/0.27	226 (1/20) 0/0	303–304 (2/20) 0/0.39
Pilot Whale	265 (1/20) 0/0	243 (1/20) 0/0	161–174 (6/17) 0.82/0.79	292 (1/19) 0/0	148 (1/20) 0/0	313–317 (4/18) 0.39/0.70	216 (1/20) 0/0	223 (1/20) 0/0	307 (1/20) 0/0
Hedgehog	260–272 (5/20) 0.65/0.70	225–230 (5/20) 0/0.10	268–172 (3/20) 0.40/0.56	321–325 (2/20) 0.45/0.48	148–154 (2/20) 0.25/0.30	345 (1/20) 0/0	151–157 (2/20) 0.25/0.30	213–227 (7/20) 0.65/0.81	303 (1/20) 0/0
Shrew	309–329 (11/20) 0.80/0.88	254–256 (3/20) 0.50/0.45	221–223 (2/20) 0/0	281 (1/20) 0/0	n/a	n/a	n/a	n/a	309–313 (4/19) 0.26/0.25
Dugong	269–273 (4/20) 0.45/0.57	225 (1/17) 0.05/0.05	176 (1/20) 0/0	274 (1/19) 0/0	138 (1/17) 0/0	289 (1/18) 0/0	n/a	222 (1/17) 0/0	294–298 (3/18) 0.50/0.41
Tenrec	n/a	n/a	n/a	281 (1/20) 0/0	n/a	n/a	n/a	n/a	316–319 (4/15) 0.33/0.55
Tammar wallaby	249–291 (9/16) 0.79/0.83	n/a	193–195 (2/16) 0.06/0.06	281 (1/16) 0/0	149 (1/16) 0/0	n/a	n/a	191–293 (14/15) 0.60/0.94	325–332 (5/10) 0.20/0.70
Quoll	241–243 (2/9) 0.11/0.11/0.10	318–342 (6/8) 0.50/0.81/0.72	n/a	297 (1/20) 0/0/0	148 (1/20) 0/0/0	n/a	203 (1/14) 0/0/0	n/a	299 (1/15) 0/0/0
Platypus	245–263 (2/15) 0/0.13/0.12	346–382 (7/13) 0.85/0.72/0.64	214–226 (4/15) 0.13/0.36/0.32	n/a	145 (1/11) 0/0/0	n/a	208 (1/18) 0/0/0	n/a	298 (1/15) 0/0/0
Echidna	248–252 (4/15) 0.20/0.57	372–376 (4/13) 0/0.65	n/a	317 (1/14) 0/0	142 (1/12) 0/0	278 (1/20) 0/0	205–213 (5/14) 0.50/0.76	194–196 (2/12) 0.09/0.09	298 (1/17) 0/0

Allelic Range (number of alleles/number of individuals successfully genotyped) Observed Heterozygosity/Expected Heterozygosity.
* indicates sequenced loci.

doi:10.1371/journal.pone.0029582.t002

exhibits polymorphism through addition/removal of repeats only, has no or a non-significant fraction of homoplastic alleles, and has a simple repeat structure with mutational dynamics in line with current models of microsatellite evolution [45,46].

We carried out cross-species direct PCR sequencing of the five most successfully genotyped microsatellite loci, namely C2-1218, C2-1915, C4-1514, C9-1918 and C17-4243 (Table 2). Four homozygous allele variants (where available) were sequenced for each species. Information S8 presents an overview of these results, with total fragment length, microsatellite length and microsatellite sequence given for all variants of successfully sequenced individuals. The direct PCR sequencing success rate was average (42%), regardless of previous genotyping success. Of 56 cases

where between two to four sequences per locus per species could be retrieved, we found 26 cases of intraspecific polymorphism (i.e. length and/or sequence polymorphism), with a total of 31 new intraspecific alleles (36 if we include chimpanzee, for which we have no genotyping data). Here we define a 'new allele' as an allelic variant of an arbitrary 'ancestral' allele (Information S8).

Ten out of the 31 new allele variants showed a difference between total length change and microsatellite length change. These differences are most likely the result of short indels occurring in flanking regions. However, we cannot be certain for all cases, due to the absence of flanking sequence information, and because genotyping errors cannot be completely ruled out. In the other 21 comparisons (68%), changes in total locus length were consistent

with repeat addition or removal in the microsatellite sequence. Although six cases of size homoplasy were observed (identical size, different sequence), only two originated from mutations in both microsatellite and flanking sequences, the other four cases originating from a point mutation within the microsatellite sequence. Finally, in all cases, addition/removal of one or more motifs occurred in the longest pure tract(s) of dinucleotide repeats.

Discussion

Microsatellites are currently one of the most popular types of genetic markers for molecular ecology, forensics and genome mapping studies. Their evolutionary dynamics have been extensively studied [1,3,45], and new analytical approaches are continually being developed [47,48]. However, their use could be facilitated, and even extended, if microsatellite markers could be readily transferred between species. Most attempts to transfer microsatellites across species are hindered by the accumulation of point mutations in microsatellite flanking sequences and/or the decay of microsatellite sequences over time [1,12]. But the recent finding that scores of microsatellite loci are indeed conserved across vertebrate genomes [25,26] has offered new hopes of significantly increasing success rates in developing cross-species microsatellite markers than have been observed to date [12].

Here we described a novel combination of *in silico* and wet-lab approaches to develop a set of microsatellite markers with broad, potentially universal, utility across the Mammalia (Figure 1). We demonstrated that an easily adaptable and reproducible protocol can be used to extract highly conserved microsatellite loci from multiple genome alignments, design degenerate primers and implement a set of microsatellite loci across vastly distant species – in this case 18 mammalian species that shared a common ancestor no earlier than 160 million years ago [24]. Although there are anecdotal reports of exceptional conservation in other taxa, e.g. turtles [49] and fish [50], this extensive transferability exceeds that of any prior cross-species study in mammals, and thus radically alters the conventional assumption that cross-species amplification of microsatellite loci is limited to closely related species [12].

Focusing our analysis on the entire breadth of the Mammalia (eutherians, marsupials and monotremes) ensured a large evolutionary scope as well as a solid genomic framework where scores of conserved microsatellites have been identified [26]. Given the extensive species divergence, it was expected that only a small fraction, if any, of the subset of widely conserved loci would not only provide a substrate to develop mammal-wide PCR primers, but also contain a polymorphic microsatellite sequence in all genomes. Our investigation shows that contrary to this common expectation, mammalian genomes contain a significant number of potential mammal-wide microsatellite markers. First, the proportion of microsatellite loci found to contain potential conserved primer sites in a first non-stringent *in silico* scan was fairly high, with conserved primer sites identified in 7.3% of the random subset of ~1,000 conserved dinucleotide repeats that we drew from our total pool of 5,596 highly conserved microsatellite loci. From those 1,000 conserved dinucleotide repeats, 19 or 1.9% were suitable to design mammal-wide degenerate primers using our stringent set of criteria (Table 1), result we view as remarkable considering the breadth of the Mammalia and the limited number of sites that we studied. Indeed, our initial subset represented only a fraction, less than a fifth, of all the microsatellites found that could be examined to identify cross-species microsatellite markers. In addition, using a more comprehensive dataset of conserved mammalian microsatellites [26], we were able to find 4,084 human dinucleotide repeats among 10,267 conserved microsatel-

lites in five non-primate mammals. Thus, by extrapolation at least 80 loci should be suitable for primer design using these selection criteria, and we anticipate that more should be identified under less stringent conditions (e.g. conservation in human-mouse-opossum). Moreover, other types than dinucleotide repeats can also be used for cross-species transfer of microsatellite markers, e.g. tetranucleotide markers, which are conserved in equivalent numbers in mammalian genomes [26]. Furthermore, if there is success in designing comparative primers useful across the Mammalia, then many more are expected to be developed from more specific comparisons, i.e. within subgroups of the Mammalia, especially with further genomic resources being acquired [27].

There was no particular relationship between PCR success and either G+C-content of PCR priming sites, genomic location, and number of degeneracy in primer sequences (Information S3). Of 19 designed primer pairs, nine were successfully optimized for mammal-wide amplification, and five were suitable for genotyping and sequencing. A number of methodological choices were made to decrease costs, but they may have reduced success rates in genotyping and sequencing, e.g. Chelex extraction method (impure DNA extract), M13-genotyping (primer dimers, inconsistent fluorescent signal), use of degenerate primers (low amplification), and direct PCR sequencing (low quality reads). We would expect a significant increase in success rate using clean extraction methods (extraction kit, phenol-chloroform protocol), standard fluorescent genotyping, non-degenerate primers and clone sequencing. In addition, we had little or no control on sampling and DNA quality for most of our samples, which may have had detrimental consequences on the overall quality of our results. For example, low polymorphism in rats and pilot whales could be explained by our samples originating from inbred populations [51] and pod strandings, respectively [52]. Drawing on these experiences, guidelines are outlined in the supplementary materials to help others planning to use conserved microsatellites to develop comparative primers (Information S9).

Overall, our cross-species primers still yielded good genotyping results for five of the nine fully optimized loci. Intraspecies polymorphism was strongly associated with length and purity of repeat tracts, which emphasized the importance of examining the sequence structure of microsatellites to select polymorphic genetic markers. Sequence information demonstrated that most changes (68%) in total fragment length at the five loci were attributable to mutations in the microsatellite sequence rather than in the flanking sequences, suggesting that cross-species primers designed for these loci are invaluable candidates for being employed as universal genetic markers across the Mammalia, as it has already been demonstrated for the under-studied short-beaked echidna [8].

Our findings establish a new paradigm in that they demonstrate that with the emergence of large numbers of genome sequences for a given taxonomic group, universal sets of microsatellite markers can be generated for that group, using a simple protocol. Provided that such sets are fully characterized and tested for confounding influences in the the different species of interest (e.g. linkage and deviations from the Hardy-Weinberg equilibrium), and standardized for use in different laboratories, this creates the genuine possibility of developing large panels of microsatellites with cross-species transferability and known genomic context [16], enabling true inter-study comparability that have long been sought but never before obtained.

Supporting Information

Information S1 Nature and origin of mammalian samples.
(PDF)

Information S2 28-way alignment of conserved microsatellites. Both flanking sequences of the conserved microsatellite (represented by the black bar) contain a stretch of orthologous sequences potentially suitable to design cross-species primers (indicated with orange boxes). Locus C2-1218 at chr2:17,699,950–17,700,450 (UCSC hg18).

(PDF)

Information S3 Description of 19 primer pairs selected for PCR optimization.

(PDF)

Information S4 Distribution of human microsatellites conserved in nine non-primate species. The human data (in bold) correspond to the total number of human microsatellites found to be conserved in at least one species. Species-specific subsets correspond to the number of human microsatellites that are conserved in at least each one of those species. Numbers in brackets indicate numbers for the whole genome (excluding Y chromosome).

(PDF)

Information S5 Electrophoresis of PCR products for C2-1218 in 17 mammals. *Bta*: cow, *Oar*: sheep, *Cfa*: dog, *Mmu*: mouse, *Fca*: cat, *Mex*: tamar wallaby, *Oan*: platypus, *Gme*: pilot whale, *Dma*: quoll, *Rno*: rat, *Ddu*: dugong, *Sar*: shrew, *Eeu*: hedgehog, *Tac*: echidna, *Hsa*: human, *Tad*: dolphin, *Ete*: tenrec. (negative control: water).

(PDF)

Information S6 Polymorphic Information Content (PIC) at nine cross-species microsatellite loci based on at most 20 individuals per species.

(PDF)

Information S7 UCSC 28-way alignment of five cross-species microsatellite locus showing species of interest (in order: C2-1218; C2-6868; C2-1915; C4-1514; C17-4243). (A) Flanking sequences. Underscores represent the

microsatellite sequence; positions are counted upstream and downstream from the microsatellite. Boxes indicate primer sites, dashes gaps and dots bases identical to human. (B) Microsatellite sequence. Array length is shown in brackets. UCSC assemblies: Human (hg18), chimp (panTro2), mouse (mm8), rat (rn4), cow (bosTau3), dog (canFam2), cat (felCat3), shrew (sorAra1), hedgehog (eriEur1), armadillo (dasNov1), elephant (loxAfr1), tenrec (echTell), opossum (monDom4), platypus (ornAna1).

(PDF)

Information S8 Allele length, and microsatellite length and sequence variation in at most four individuals for five cross-species microsatellite loci. L_{allele} : allele length (bp); L_{micro} : microsatellite length (bp); an individual (Indiv 1) 1 is used as a reference to measure length variation (+/–).

(PDF)

Information S9 Guidelines to facilitate the identification, design, optimization and implementation of comparative microsatellite primers.

(PDF)

Acknowledgments

We warmly thank the many individuals and institutes that have generously contributed to our collection of mammalian samples: J. Hickford, L. Møller, M. Oremus and S. Baker, I. Vargas-Jentsch, M. Hale, G. Yannic and J. Hausser, D. Tautz, B. Robertson, A. Amanzadeh and F. Shokri, A. Stone, S. Goodman, A. MacMahon and D. Blair, J. Graves, M. Cardoso, C. Whittington, S. Nicol, and P. Rismiler. Names of affiliated institutes can be found in Information S1. We are also grateful to Luz-Andrea Pfister for sequencing the chimpanzee samples.

Author Contributions

Conceived and designed the experiments: EB NJG. Performed the experiments: EB JSB. Analyzed the data: EB JSB. Contributed reagents/materials/analysis tools: NJG. Wrote the paper: EB.

References

- Buschiazzi E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* 28: 1040–1050.
- Hannan AJ (2010) Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for ‘missing heritability’. *Trends in Genetics* 26: 59–65.
- Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* 5: 63–69.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: A review. *Molecular Ecology* 11: 1–16.
- Raveendran M, Harris RA, Milosavljevic A, Johnson Z, Shelledy W, et al. (2006) Designing new microsatellite markers for linkage and population genetic analyses in rhesus macaques and other nonhuman primates. *Genomics* 88: 706–710.
- Raveendran M, Tardif S, Ross CN, Austad SN, Harris RA, et al. (2008) Polymorphic microsatellite loci for the common marmoset (*Callithrix jacchus*) designed using a cost- and time-efficient method. *American Journal of Primatology* 70: 906–910.
- Vanpé C, Buschiazzi E, Abdelkrim J, Morrow G, Nicol S, et al. (2009) Development of microsatellite markers for the short-beaked echidna using three different approaches. *Australian Journal of Zoology* 57: 219–224.
- Abdelkrim J, Robertson BC, Stanton JAL, Gemmell NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques* 46: 185–192.
- Primmer CR, Møller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* 5: 365–378.
- Gemmell NJ, Allen PJ, Goodman SJ, Reed JZ (1997) Interspecific microsatellite markers for the study of pinniped populations. *Molecular Ecology* 6: 661–666.
- Barbara T, Palma-Silva C, Paggi GM, Bered F, Fay MF, et al. (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759–3767.
- Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392: 917–920.
- Primmer CR, Painter JN, Koskinen MT, Palo JU, Merila J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology* 36: 348–360.
- 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.
- Osborne AJ, Brauning R, Schultz JK, Kennedy MA, Slate J, et al. (2011) Development of a predicted physical map of microsatellite locus positions for pinnipeds, with wider applicability to the Carnivora. *Molecular Ecology Resources* 11: 503–513.
- Stallings RL (1995) Conservation and evolution of (CT)_n/(GA)_n microsatellite sequences at orthologous positions in diverse mammalian genomes. *Genomics* 25: 107–113.
- Farber CR, Medrano JF (2004) Identification of putative homology between horse microsatellite flanking sequences and cross-species ESTs, mRNAs and genomic sequences. *Animal Genetics* 35: 28–33.
- Liewlaksanecyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical Applied Genetics* 109: 361–369.
- Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, et al. (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science* 168: 195–202.
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* 99: 125–132.
- Dawson DA, Horsburgh GJ, Küpper C, Stewart IRK, Ball AD, et al. (2009) 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.
- Küpper C, Burke T, Szekely T, Dawson D (2008) Enhanced cross-species utility of conserved microsatellite markers in shorebirds. *BMC Genomics* 9: 502.

24. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453: 175–183.
25. Buschiazzo E, Gemmell NJ (2009) Evolutionary and phylogenetic significance of platypus microsatellites conserved in mammalian and other vertebrate genomes. *Australian Journal of Zoology* 57: 175–184.
26. Buschiazzo E, Gemmell NJ (2010) Conservation of human microsatellites across 450 million years of evolution. *Genome Biology and Evolution* 2010: . pp 153–165.
27. Genome 10K Community of Scientists (2009) Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *Journal of Heredity* 100: 659–674.
28. Walsh PW, Metzger DA, Higushi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10: 506–513.
29. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, et al. (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Research* 39: D876–D882.
30. La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23.
31. Goecks J, Nekrutenko A, Taylor J, The Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 11: R86.
32. Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research* 18: 30–38.
33. Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, et al. (2007) 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Research* 17: 1797–1808.
34. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
35. Gadberry MD, Malcomber ST, Doust AN, Kellogg EA (2005) Primaclade—a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* 21: 1263–1264.
36. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132: 365–386.
37. Moore SS, Hale P, Byrne K (1998) NCAM: a polymorphic microsatellite locus conserved across eutherian mammal species. *Animal Genetics* 29: 33–36.
38. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.
39. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.
40. Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639–655.
41. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.
42. Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* 140: 679–695.
43. Garza JC, Desmarais E (2000) Derivation of a simple microsatellite locus from a compound ancestor in the genus *Mus*. *Mammalian Genome* 11: 1117–1122.
44. Estoup A, Jarne P, Cornuet JM (2002) Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591–1604.
45. Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435–445.
46. Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: Insights from a comparison of humans and chimpanzees. *Genetics* 168: 383–395.
47. Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the MSVAR method. *Genetics*;doi:10.1534/genetics.1110.121764.
48. Wu C-H, Drummond AJ (2011) Joint inference of microsatellite mutation models, population history and genealogies using trans-dimensional MCMC. *Genetics*;doi:10.1534/genetics.1110.125260.
49. FitzSimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* 12: 432–440.
50. Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London Series B: Biological Sciences* 263: 549–557.
51. Robertson BC, Gemmell NJ (2004) Defining eradication units to control invasive pests. *Journal of Applied Ecology* 41: 1042–1048.
52. Amos B, Schlötterer C, Tautz D (1993) Social structure of pilot whales revealed by analytical DNA profiling. *Science* 260: 670–672.