Supporting Information

Rosenblum et al. 10.1073/pnas.1300130110

SI Methods

Isolate Selection and Genomic Sequencing. We selected 29 *Batrachochytrium dendrobatidis* (Bd) isolates for de novo sequencing with multiple samples from the focal regions. We focused on the Americas, with 7 isolates from the western United States, 6 isolates from the eastern United States and Canada, and 13 isolates from Latin America. We also included isolates from Japan, Australia, and South Africa. The chytrid, *Homolaphlyctis polyrhiza* (Hp; JEL142), which we sequenced previously and does not infect frog skin (1), served as an outgroup to root the evolutionary analyses. Each Bd culture was grown at room temperature for 7–14 d on 1% tryptone and 1% agar plates. Zoospores were flooded from plates and concentrated with a tabletop centrifuge. We extracted genomic DNA using a modified protocol from ref. 2 with 2% (vol/vol) SDS as the extraction buffer or the Qiagen DNeasy Kit. We used the Illumina GA IIx and Illumina HiSeq platforms at the Cornell Core Laboratories. We obtained an average of 24× sequencing depth per isolate (Table S1). All genomic data are accessioned in the National Center for Biotechnology Information Short Read Archive (accession no. SRA062886).

Sequence Alignment and SNP Calling. Illumina reads were quality-filtered (60% bases phred score > 20), low-quality ends were trimmed, and reads were aligned to the genome of JEL423 (ver 17-Jan-2007) with Stampy (3) applying base call recalibration. Best practice protocol for variant calling in GATK was applied (4). Duplicate reads were marked with Picard, and reads containing indels were realigned with Smith–Waterman (GATK walkers RealignerTargetCreator; IndelRealigner). Final variant calls were made and filtered of false positives with GATK UnifiedGenotyper and VariantFiltration walkers. From the resulting SNP dataset, we generated a stringent dataset for downstream analyses by applying a minimum 10× coverage filter on a strain-by-strain basis. Genotype calls with low coverage were converted to missing data calls.

Meta-Analysis. For a subset of comparative analyses, we integrated our data with previously published data from an additional 20 Bd isolates (5). The isolate selection from the two datasets was geographically complementary [our sampling focused on the New World, and the work by Farrer et al. (5) focused on Europe]; 20 isolates from the study by Farrer et al. (5) were sequenced using SOLiD technology. We aligned the SOLiD reads to the Bd genome first using BFAST aligner (6) followed by applying the same variant calling protocol as used with the Illumina data. The read depth was found to be significantly lower in the SOLiD dataset compared with our Illumina data. To reduce false-positive counts, the final SNP calls were made by only considering sites called as variable in the Illumina dataset. De novo SNP calling in SOLiD data significantly increased the number of false-positive SNPs, and subsequent multidimensional scaling (MDS) plots showed separation of strains purely by sequencing technology. However, recalling SNPs with high-depth stringencies based first on our Illumina dataset removed these biases.

Phylogenetic Reconstruction. We estimated rooted phylogenies separately for our 29 isolates and the 49 isolates in the meta-analysis using Hp as the outgroup. Hp alleles were called from whole-genome alignment of Hp and JEL423 genome using the tool Mercator (7) and Prank (8) extracting positions defined as SNPs based on the JEL423 reference position with custom Perl scripts built with BioPerl (9). We generated trees using 101,931 sites in the 29-isolate analysis and 76,515 sites in the 49-isolate analysis, which contained informative SNPs from the nuclear genome. We used the parsimony criterion in PAUP* 4.0 to reconstruct the evolutionary history with these unphased nuclear SNPs (10). We searched tree space by performing 100 search replicates using tree–bisection–reconnection to swap branches. The SNPs were encoded to distinguish six character states. The first three character states (0–2) were used for SNPs, where at least one Bd isolate shared a common allele with Hp (homozygous with respect to Hp allele, and heterozygous and homozygous with respect to an alternate allele). The next three character states (3–5) were used for SNPs, where Bd strains and Hp did not have an allele in common, and they also distinguished the three possible genotypes. We used a hetequal character transition matrix as previously described (11). We performed 200 bootstrap replicates to generate node support values under the parsimony optimality criterion.

Loss of Heterozygosity Analysis. We used a hidden Markov model (HMM) to identify regions in the genome with long stretches of homozygosity. We analyzed the 16 largest supercontigs, which accounted for 98.6% of the total SNP dataset. The HMM was constructed using the RHmm R package (12). We used a non-overlapping sliding window approach, where the number of heterozygous sites was calculated in each 100-bp window and fit to a model for each supercontig by implementing the Baum–Welch algorithm. Then, we used the Viterbi algorithm to predict loss of heterozygosity (LOH) regions with the fitted HMM and the observed SNP data. We cleaned the raw predicted LOH calls by filtering out very short (<1 kbp) regions and obvious false positives, which were likely caused by low read coverage.

Divergence Estimation. We used BEAST (v 1.7.3; 40) to sample from the posterior density of time-calibrated trees for 49 isolates of Bd with Hp as the outgroup. We applied a constant-size coalescent tree prior and a strict molecular clock with a rate of 0.0081 substitutions per site per million years, a rate that has been used in previous studies of fungi (13, 14). Our phylogenomic dataset was unpartitioned and 23,597,406 sites in length, a nearly comprehensive sample of the nuclear genome. We assumed that sites not identified as SNPs were invariant for the nucleotide recorded for our reference strain as above. We allowed the Markov chain Monte Carlo (MCMC) sampling to burn in for the first 10 million generations. After the burn-in period, we sampled every 1,000 generations for an additional 20 million generations. Preliminary analyses invoking a relaxed clock or an exponentially increasing population under the coalescent tree prior were quantitatively similar to our reported results. We also conducted divergence–time analyses using datasets subsampled at the level of the supercontig. To the greatest possible extent, we matched region lengths for LOH and non-LOH segments along the same supercontig. BEAST was used for tree estimation using 30 subsampled datasets. Chain lengths were 10 million generations, the first one-half of which were discarded as burn in. Other details follow the previous description of our BEAST analysis with the concatenated dataset. Using the maximum clade credibility trees for each segment, we compared estimated heights of the global panzootic lineage (GPL) between paired LOH and non-LOH regions of the same supercontig. We assumed that divergence dates were log-normal in distribution, and thus, statistical comparison between data partitions used log-transformed dates.

Chromosome and Copy Number Variation. Ploidy and aneuploidy were estimated using SNP read depth and allele frequencies. The mean depth across SNPs for each supercontig was extracted from the variant call format (VCF) file, and the estimated chromosome copy number was determined by identifying clusters of supercontig depths of similar value using *k*-means clustering with the pamk function of the R package fpc. Each cluster was assigned to one of the following copy numbers using the distribution of all SNP allele frequencies for that supercontig: monosomy, disomy, trisomy, or tetrasomy. The expectation was that allele frequencies would have a unimodal distribution centered at 0.5 for disomic chromosomes, a bimodal distribution of 0.67 and 0.33 for trisomic chromosomes, and a trimodal distribution of 0.25, 0.5, and 0.75 for tetrasomic chromosomes. Coverage per base was extracted using GATK DepthofCoverage walker. We detected copy number variation as significant expansion or loss of regions of DNA in each strain using a Bayesian approach implemented in the R package cn.mops (15). Read counts were extracted from BAM files mapped onto the reference sequence and specifically normalized for each isolate using the chromosomal numbers estimated during aneuploidy detection. We used a 1,000-bp minimum window size for a copy number variation event, a lower threshold for detection at −1 (corresponding to a twofold reduction in copy number), and an upper detection limit of 1.

Tests for Selection. We scanned for selection by implementing the sequence divergence test dN/dS using JEL423 gene models as reference. We used the Yang and Nielsen method, yn00, in PAML 4.6 (16) to calculate pairwise dN/dS for each gene for comparisons between the GPL isolates ($n = 26$) and the basal UM142 isolate. We then conducted an analysis of Gene Ontology (GO) enrichment function for the genes showing elevated nonsynonymous SNP counts. Enrichment significance was calculated with the GOstats package (17) using GO to gene assignments generated by an InterproScan version 5–RC3 (https://code.google.com/p/interproscan/) (18) analysis of the *Bd* proteome. Analysis scripts and datasets are available from https://github.com/stajichlab/bd_popgen.

1. Joneson S, Stajich JE, Shiu SH, Rosenblum EB (2011) Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathog* 7(11):e1002338.
2. Zolan ME, Pukkila PJ (1986) Inheritance of DNA methylation in *Coprinus cinereus*. *Mol Cell Biol* 6(1):195–200.
3. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21(6):936–939.
4. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9): 1297–1303.
5. Farrer RA, et al. (2011) Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proc Natl Acad Sci USA* 108(46):18732–18736.
6. Homer N, Merriman B, Nelson SF (2009) BFAST: An alignment tool for large scale genome resequencing. *PLoS One* 4(11):e7767.
7. Dewey CN (2012) Whole-genome alignment. *Methods Mol Biol* 855:237–257.
8. Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* 102(30):10557–10562.
9. Stajich JE, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 12(10):1611–1618.
10. Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4* (Sinauer, Sunderland, MA).
11. James TY, et al. (2009) Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. *PLoS Pathog* 5(5):e1000458.
12. Taramasco O, Bauer S (2012) *Hidden Markov Models Simulations and Estimations*. Available at http://r-forge.r-project.org/projects/rhmm/. Accessed December 31, 2012.
13. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155.
14. Joneson S, Stajich JE, Shiu SH, Rosenblum EB (2011) Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathog* 7(11):e1002338.
15. Klambauer G, et al. (2012) cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res* 40(9):e69.
16. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
17. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23(2):257–258.
18. Hunter S, et al. (2012) InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Res* 40(Database issue):D306–D312.

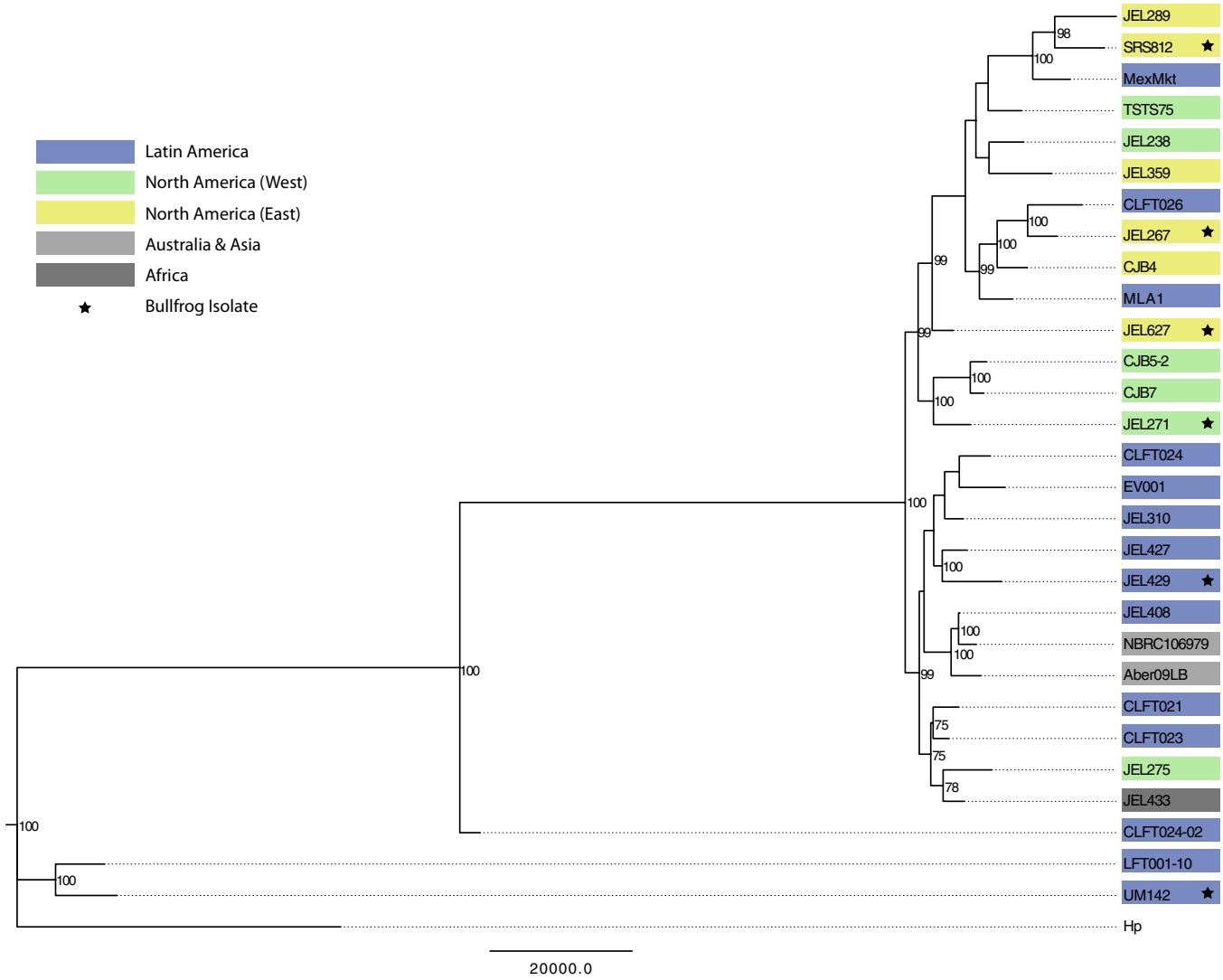
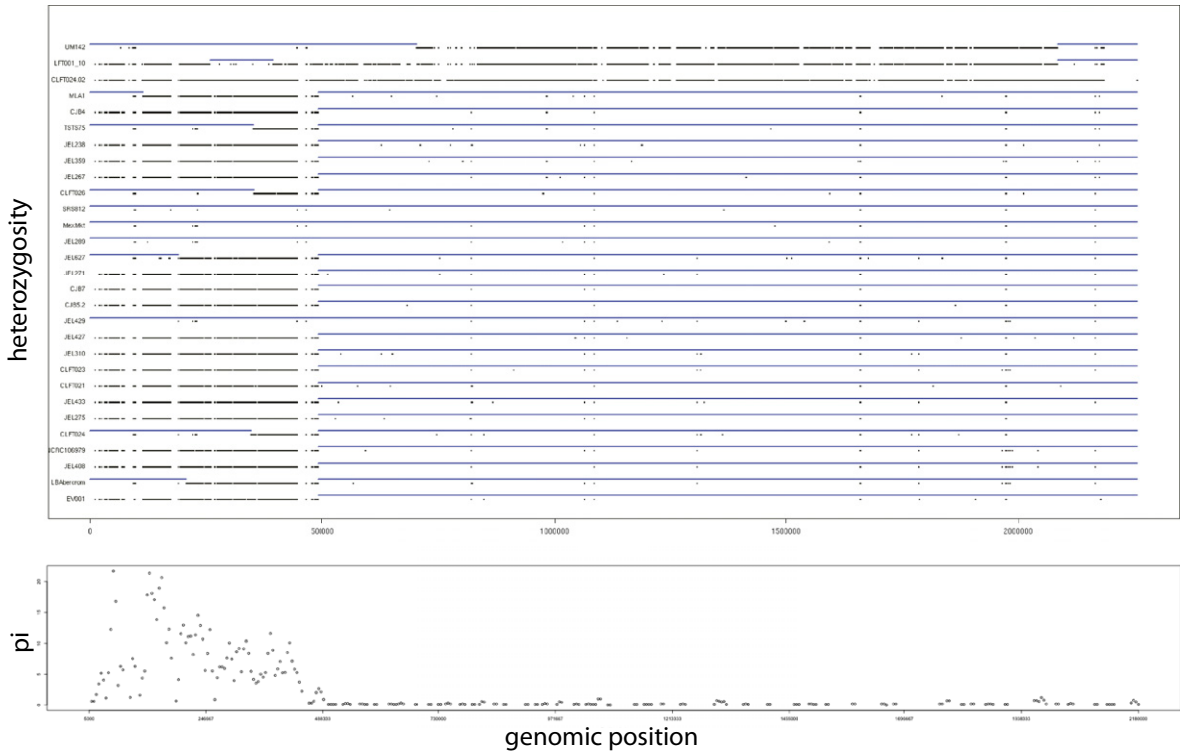


Fig. S1. Rooted Bd phylogeny for 29 isolates that were the focus of this study based on 101,931 SNPs.

A. Supercontig 2



B. Supercontig 7

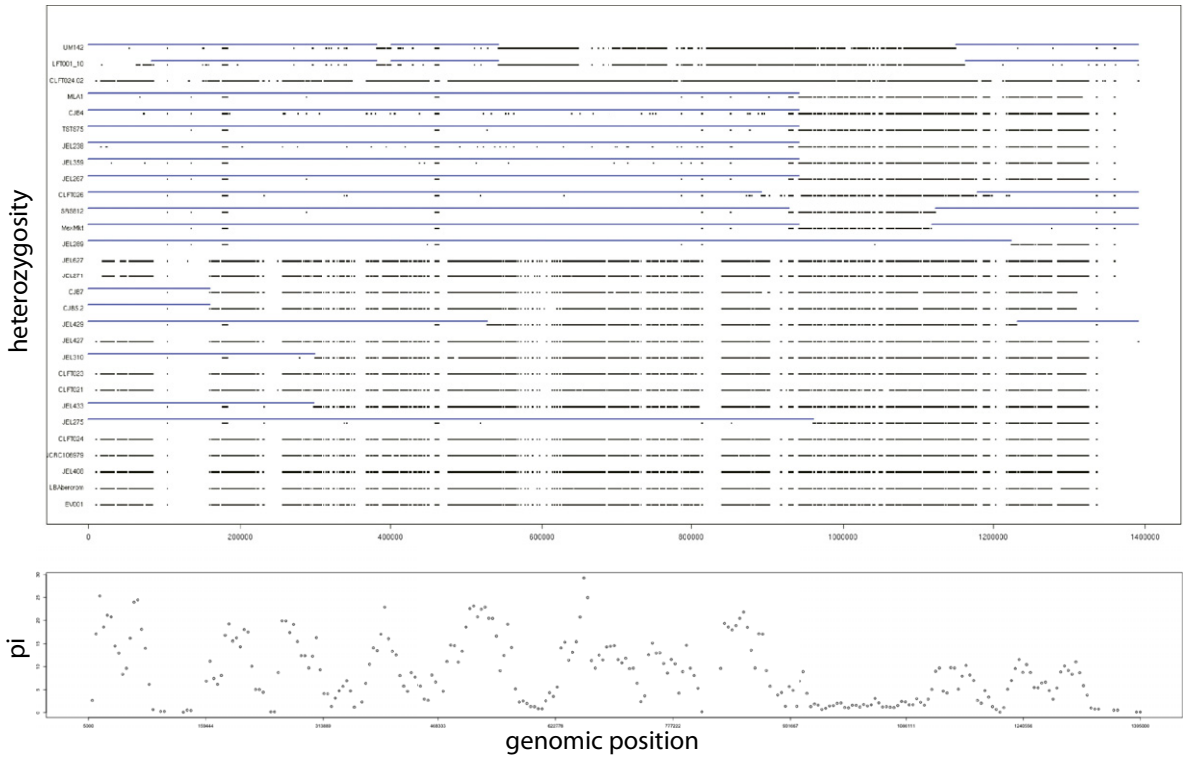


Fig. S2. Examples of shared LOH events on (*A*) supercontig 2 and (*B*) supercontig 7. *Upper* shows heterozygous positions (black dots) and LOH regions (blue lines) for each isolate individually. *Lower* shows average nucleotide diversity, π, in sliding windows (10,000 bp) across the same chromosomal segment for 26 GPL isolates.

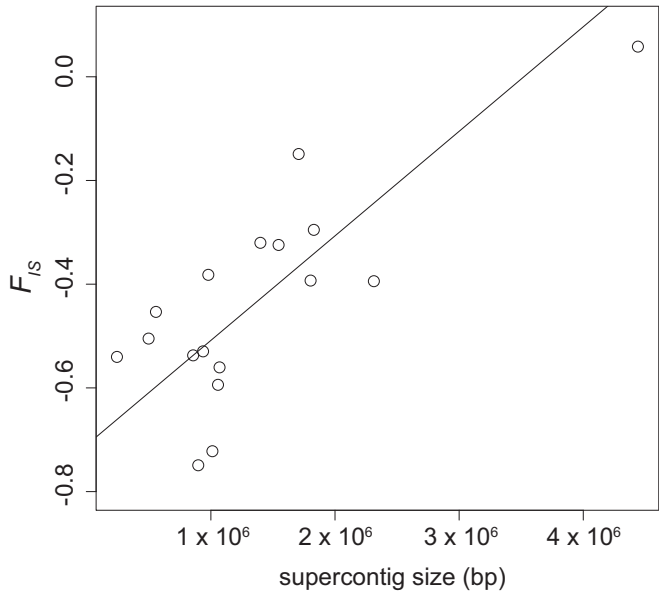


Fig. S3. Scatter plot of mean F_{IS} for each of the 17 largest supercontigs by supercontig size. A best-fit linear regression is shown ($r^2 = 0.70$; $P < 0.01$).

Table S1. Bd isolates included in the resequencing study

Sample Identification	Collection locality	Amphibian host	Sequencing depth
CJB4	Yosemite National Park, CA	*Rana muscosa/sierrae*	15
CJB5-2	Sierra National Forest, CA	*Rana muscosa/sierrae*	14
CJB7	Kings Canyon National Park, CA	*Rana muscosa/sierrae*	17
CLFT021	Serra do Japí, Brazil	Unidentified tadpole	8
CLFT023	Monte Verde, Brazil	*Hypsiboas* sp.	17
CLFT024	Estrada da Graciosa, Brazil	*Hylodes cardosoi*	34
CLFT024-02	Estrada da Graciosa, Brazil	*Hylodes cardosoi*	38
CLFT026	Reserva Betary, Brazil	*Hypsiboas faber*	28
EV001	Ubaque, Colombia	*Rheobates palmatus*	25
JEL238	Mesquite Wash, AZ	*Lithobates yavapaiensis*	16
JEL267	Mont-Saint-Hilaire, Quebec, Canada	*Lithobates catesbeianus*	32
JEL271	Point Reyes, CA	*Lithobates catesbeianus*	29
JEL275	Clear Creek Co., CO	*Anaxyrus boreas*	38
JEL289	Milford, ME	*Lithobates pipiens*	17
JEL310	Fortuna, Panama	*Smilisca phaeota*	18
JEL359	Berlin, NH	*Lithobates clamitans*	14
JEL408	El Cope, Panama	*Colostethus inguinalis*	31
JEL427	El Yunque, Puerto Rico	*Eleutherodactylus coqui*	17
JEL429	Merida, Venezuela	*Lithobates catesbeianus*	32
JEL433	Namaqualand, South Africa	*Xenopus laevis*	31
JEL627	Bethel, ME	*Lithobates catesbeianus*	15
Aber09LB	Abercrombie River, Australia	*Litoria booroolongensis*	26
LFT001-10	Serra do Japí, Brazil	*Hylodes ornatus*	48
MexMkt	Mercado Emilio Carranza, Mexico City	*Hyla eximia*	24
MLA1	Las Higueritas Natural Reserve, Argentina	*Hypsiboas cordobae*	33
NBRC106979	Chuo-ku, Japan	*Ceratophrys cranwelli*	18
SRS812	Savanna River, SC	*Lithobates catesbeianus*	17
TST75	Yosemite National Park, CA	*Rana muscosa/sierrae*	14
UM142	Ypsilanti, MI (market)	*Lithobates catesbeianus*	32