Supporting Information

Song *et al.* 10.1073/pnas.0803076105

SI Methods

Numt Amplification. For both grasshoppers and crayfish, the genomic DNA was extracted by using Qiagen DNeasy extraction protocols for animal tissue. Two different approaches of numt generation were made between the two organisms. For grasshoppers, we performed PCR by using high fidelity Elongase polymerase (Invitrogen) and conserved primers on genomic DNA extract from single individuals of each species and cloned the PCR products by using the TOPO Cloning Kit (Invitrogen). A 488-bp fragment of COI was amplified by using a pair of insect conserved primers described in Simon *et al.* (1), C1-J-1751 (Ron) and C1-N-2191 (Nancy). This amplified region completely overlapped with the Folmer region (2) of COI. To test whether the same classes of numts were generated with both primer pairs, we also amplified a 709-bp fragment of COI gene from *S. americana* by using the Folmer primers, which yielded 19 paralogous haplotypes with in-frame stop codons, the mean sequence divergence from the mtDNA ortholog being 3.89%. For crayfish, we used the Folmer primers to generate 709 bp of the COI gene from genomic DNA extract. For 93 individuals, we were able to obtain clean COI sequences; however, 79 individuals from southern populations of *O. australis* and *O. barri* yielded ambiguous sequences. We therefore cloned 15 individuals from those populations and also cloned eight individuals (including *O. incomptus* and *O. packardi*) that previously yielded clean unambiguous COI sequences to determine whether numts were present but not being detected without cloning. A total of 23 crayfish individuals from four species were cloned. PCR condition was as following: 10–12 min of hot start at 94–95°C, followed by 35–40 cycles of 94°C for 1 min, 40–45°C for 1 min, and 72°C for 1 min, and a 7-min final extension at 72°C. The resulting PCR products were cloned (125 clones for grasshoppers, 560 for crayfish), which were amplified by PCR and sequenced by using M13 primers (Table S3). Sequencing reactions were performed by using ABI BigDye terminator sequencing technology and raw sequence files were edited in Sequencher 4.6 (GeneCodes Corp.). Contigs of edited sequences and primer sequences were created to identify the haplotypes and both ends of the sequences matching the primer sequences were excised to remove artificial nucleotide similarity introduced by PCR amplification, resulting in the final lengths of 439 bp for grasshoppers and 658 bp for crayfish.

1. Simon C, *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* 87:651–701.

2. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3:294–299.

Table S1. Summary of sequence characteristics of numt haplotypes generated in this study

Species	Unique haplotypes	mtDNA AT%	Numt haplotypes	% numt	Numt AT%	Length variation	Deletion	Insertion	Point mutation	Stop codon
Grasshoppers										
Locusta migratoria ($n = 1$)	14	65.83	3	20.00	65.74	379–437	1–7	0–2	31–76	0–2
Acrida willemsei ($n = 1$)	13	65.15	6	46.15	65.44	356–440	0–3	0–1	6–61	0–17
Calliptamus italicus ($n = 1$)	10	62.19	2	20.00	61.57	438–439	0–1	0	1–18	0–14
Schistocerca americana ($n = 1$)	25	61.73	11	44.00	60.88	438–449	0–4	0–1	28–115	0–11
Crayfish										
Orconectes australis ($n = 79$)	122	63.12	60	49.18	63.40	654–659	0–4	0–10	11–81	0–24
Orconectes barri ($n = 16$)	60	63.08	46	76.67	63.66	646–659	0–7	0–2	4–86	0–24
Orconectes incomptus ($n = 4$)	4	63.11	1	25.00	63.01	657	1	0	1	17
Orconectes packardi ($n = 20$)	24	63.09	4	16.67	63.13	657	1	0	52–81	19–22

Unique haplotypes were determined by comparing each haplotype against each other and against orthologous mtDNA sequences. Numt haplotypes were identified by indels, in-frame stop codons, compositional bias, and phylogenetic analyses. % numt represents the proportion of numt haplotypes to unique haplotypes within a given species. The mean sequence divergence of haplotypes from the mtDNA orthologs was different for each species of grasshoppers (*C. italicus* [0.63%], *L. migratoria* [3.08%], *A. willemsei* [5.62%] and *S. americana* [7.89%]) and was generally high for crayfish from all 11 localities ranging from 9.32 to 12.15%.

Table S2. Population-specific variation in the number of numt haplotypes in four crayfish species. AL: Alabama; TN: Tennessee; TNAL: Tennessee south of Grundy County to Alabama border; KY: Kentucky; KYTN: Kentucky-Tennessee border

Species	State	Population	Unique haplotypes	Numt haplotypes	Mean sequence divergence (%) from mtDNA ortholog (range)
O. australis	AL	LIMROCK	13	12	10.05 (1.71–11.59)
	AL	DOUGGREEN	1	0	N/A
	AL	HERING	7	7	13.12 (12.88–13.46)
	AL	SCOTT	1	1	12.67
	AL	SHELTA	22	20	12.32 (7–14.02)
	TN	BAILEY	2	0	N/A
	TN	BLINDFISH	3	0	N/A
	TN	BLOWING	1	0	N/A
	TN	CAMPSGULF	1	0	N/A
	TN	CAPSHAW	4	0	N/A
	TN	CUMBER	3	0	N/A
	TN	FALLEN	1	0	N/A
	TN	GLENCORA	1	0	N/A
	TN	JACO	1	0	N/A
	TN	LAUREL	3	0	N/A
	TN	LOSTCOVE	1	0	N/A
	TN	LOSTCREEK	1	0	N/A
	TN	LOTTDEAN	2	0	N/A
	TN	MERRY	9	0	N/A
	TN	MILLHOLLOW	2	0	N/A
	TN	NORTON	1	0	N/A
	TN	ROCKY	6	0	N/A
	TN	RUMBLING	8	7	9.99 (6.26–12.1)
	TN	SKILLMAN	1	0	N/A
	TN	VIRGIN	1	0	N/A
	TN	WATERFALL	2	0	N/A
	TN	WINCHING	2	0	N/A
	TNAL	DRY	1	0	N/A
	TNAL	FLOORLESS	1	0	N/A
	TNAL	LUSK	3	0	N/A
	TNAL	PEARSON	11	10	11.97 (11.38–12.31)
	TNAL	REDTRILLIUM	3	3	11.92 (11.72–12.11)
	TNAL	WITHER	2	0	N/A
	TNAL	WOODLEE	1	0	N/A
O. barri	KYTN	CLINTON	36	33	10.73 (0.63–12.2)
	KYTN	CORNSTARCH	16	13	10.15 (5.01–14.62)
	KYTN	GRAYSON	1	0	N/A
	KYTN	REDBUD	1	0	N/A
	KYTN	REDMOND	1	0	N/A
	KYTN	STREAM	2	0	N/A
	KYTN	TONAYS	3	0	N/A
O. incomptus	TN	CHERRY	1	0	N/A
	TN	FLYNN	1	0	N/A
	TN	NORTHFORK	2	1	0.16
O. packardi	KY	BIGSINK	1	0	N/A
	KY	CEDARCREEK	1	0	N/A
	KY	CORAL	1	0	N/A
	KY	DAVES	1	0	N/A
	KY	DUVALTS	1	0	N/A
	KY	DYKES	1	0	N/A
	KY	FLETCHER	1	0	N/A
	KY	HAIL	2	0	N/A
	KY	JUGORNOT	6	0	N/A
	KY	PINEHILL	1	0	N/A
	KY	SLOANS	2	0	N/A
	KY	TEAMERS	6	4	11.23 (8.48–13.76)

Table S3. Number of individuals used to generate numt haplotypes via cloning

Species	Cave locality	Number of Individuals*	Number of clones	Number of numt seqs†	Number of COI seqs†
O. australis	Limrock Blowing	3	77	42	3
	Larkin	1	8	4	1
	Red Trillium	1	3	1	0
	Hering	1	22	9	0
	Doug Green	1	24	11	0
	Pearson	1	26	9	13
	Shelta	2	53	39	6
	Rumbling Falls	1	39	8	23
	Subtotal	11	252	123	46
O. packardi	Teamers	1	31	7	20
	Jugornot	1	14	0	8
	Hail	2	35	0	23
	Subtotal	4	80	7	51
O. incomptus	North Fork	2	39	3	29
	Cherry	2	43	0	26
	Subtotal	4	82	3	55
O. barri	Clinton	3	112	71	9
	Cornstarch	1	34	21	5
	Subtotal	4	146	92	14
TOTAL	23	560	225	166	

*Total number of individuals does not include the 18 individuals for which numts were sequenced cleanly without cloning.
†Includes repeat numt and COI haplotypes.