

SC/66b/SH/18

A validated register and standard
nomenclature for mtDNA haplotypes of
southern hemisphere humpback whales

Debbie Steel, Carlos Olavarria, Marcia Engel, Howard
Rosenbaum, Francine Kershaw, Jen Jackson, Ana
Lucia Cypriano-Souza, Angie Sremba and C. Scott
Baker



INTERNATIONAL
WHALING COMMISSION

A validated register and standard nomenclature for mtDNA haplotypes of southern hemisphere humpback whales

Debbie Steel¹, Carlos Olavarria², Marcia Engel³, Howard Rosenbaum⁴, Francine Kershaw, Jen Jackson⁶, Ana Lúcia Cypriano-Souza^{3,7}, Angie Sremba¹ and C. Scott Baker¹

- 1) Marine Mammal Institute, Oregon State University, Newport, Oregon 97365 USA
- 2) Centro de Estudios Avanzados en Zonas Aridas, Raúl Bitrán 1305, La Serena, Chile
- 3) Projeto Baleia Jubarte/Instituto Baleia Jubarte, 45900-000 Caravelas, Bahia, Brazil
- 4) Wildlife Conservation Society, Ocean Giants Program, Global Conservation, 2300 Southern Blvd., Bronx, NY 10460, USA
- 5) Natural Resources Defense Council, 40 W 20th Street, New York, NY 1011, USA
- 6) British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
- 7) Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, 90619-900 Porto Alegre, Rio Grande do Sul, Brazil

Corresponding author: C.S. Baker
Email: scott.baker@oregonstate.edu

Summary

Here we review the published records of mtDNA control region sequences of southern hemisphere humpback whales as a first step in developing a validated register of haplotypes for future analyses of interest to the Scientific Committee of the International Whaling Commission. We restricted our initial review to sequences submitted to GenBank as ‘population datasets’ and accompanied by references to publication in peer-reviewed journals. For these 6 publications, we compared the sequence variation for all of these submissions using a standard segment of 465 base pairs in length, where available, to identify unique haplotypes, many of which were reported in more than one publication. We identify a small number of likely sequencing errors in some submissions and confirm that many haplotypes are represented by multiple GenBank accession numbers (up to 16, in some cases) and haplotype codes assigned by individual authors. From the review, we provide a ‘Rosetta stone’ to cross-reference the different internal codes and redundant GenBank codes, resolving 223 haplotypes. We also suggest a standardised nomenclature for ‘internal codes’ and GenBank codes based on precedent of publication. Although our preliminary database can be used in cross-reference with BLAST searches of GenBank for identifying known and novel haplotypes, we do not consider GenBank to be well suited for hosting a standardised register. We suggest, instead, further computational development to automate matching of haplotypes, similar to the species identification program www.DNASurveillance, and to include spatially explicit information on the location and frequencies of haplotypes, similar to the searchable database for the North Pacific humpback whales, <http://www.splashcatalog.org/mmuwildbook/>.

Introduction

Mitochondrial (mt) DNA control region sequences are now widely used by the Scientific Committee of the International Whaling Commission. In the case of the Comprehensive Assessment of southern hemisphere humpback whales, the frequencies of mtDNA haplotypes have been used extensively for analyses of population structure, population mixing and estimates of Nmin in population dynamic models (IWC 2015, Annex H).

At SC/66a it was noted that, with the growing number of publications reporting mtDNA haplotypes, there was the potential for confusion in the nomenclature used by individual researchers to refer to haplotypes and redundancy in submissions of haplotypes to GenBank, the

international repository for DNA sequence information. An intersessional working group was tasked with reviewing available information for developing a database for southern hemisphere humpback whales with a standardised nomenclature and to consider how this might be hosted by the IWC (IWC 2016, Annex H, Table3). We note a similar effort by another working group to standardize nomenclature for haplotypes of blue whales from the North Pacific and Southern Hemisphere (Lang et al. 2016)

This report summarises the first stages of the review of southern hemisphere humpback whales and the development of a ‘Rosetta Stone’ to provide cross-referencing for mtDNA haplotypes from published databases represented on GenBank. In assembling the databases and in discussion with the authors, who are represented in this report, we also detected and corrected some errors in the haplotype sequences.

Methods and Results

For this initial review, we restricted our summary to mtDNA control region sequences submitted to GenBank as population datasets (‘popsets’) and accompanied by a reference to an article in a peer-reviewed journal. With these criteria, we identified 6 datasets with accompanying references (Table 1). We also identified 3 popsets available on GenBank but lacking an accompanying reference to a peer-reviewed publication and one peer-reviewed publication with sequences submitted as Supplemental Material but not available on GenBank (Schmitt et al. 2014). Finally, we are aware of additional datasets likely to be available from the recent publication of Amaral et al. (2016) and manuscripts in review by Cypriano-Souza et al. and Rosenbaum et al. These sequences can be considered for inclusion in future reviews, depending on the outcome of subcommittee discussions.

From the 6 popset with accompanying peer-reviewed publications, we assembled a ‘Rosetta Stone’ to cross-reference the unique haplotypes (Table 2) and an accompanying Excel file with details of each dataset and the FASTA files of the GenBank submissions. The summary Excel file will be made available through the IWC OneDrive following the start of the SC/66b.

Our strategy for developing the Rosetta Stone and associated files is summarised, as follows:

1. Downloaded mtDNA control region sequences represented in 6 popsets for SH humpback whales from GenBank.
2. These are listed in the ‘Datasheets’ sheet with primary author, corresponding journal information where applicable and links back to GenBank. FASTA downloads of each dataset are included as separate sheets under first author name, including those from the 3 popsets lacking an associated reference to a publication.
3. The 6 datasets published in peer-reviewed articles were trimmed to match the popset submitted by Olavarria et al 2007 (approximately 470 base pairs in length)
 - a. This resulted in the collapse of some haplotypes within the Rosenbaum et al dataset, which were trimmed at the start.
 - b. Sequences in the Engel et al and the Valsecchi et al datasets were shorter than Olavarria et al and in some cases match more than one haplotype over the shorter length.
4. All 6 datasets were aligned with the program Sequencher to find identical sequences.
5. Matching and unique haplotypes are listed in the ‘Rosetta’ datasheet. Due to the large number of replicate sequences within the Engel et al dataset only one is represented on the ‘Rosetta’ datasheet, the rest are listed on the ‘Engel codes’ datasheet. Links between these two datasheets can be made with the accession numbers listed in column A in both sheets, such that the row labelled ‘3’ within the ‘Rosetta’ datasheet is the same haplotype as samples listed in the row labelled ‘3’ in the ‘Engel codes’ datasheet.

6. Haplotypes within the Rosenbaum et al dataset that collapsed as per 3a above are listed in the same row in the 'Rosetta' datasheet.
7. Sequences with the Engel et al dataset that match more than one haplotype as per 3b above are highlighted in purple on both the 'Rosetta' and the 'Engel codes' datasheets.
8. Previous review of the Olavarria et al dataset revealed errors with 3 haplotypes, SP51, SP56 and SP103. Once the errors were corrected, these haplotypes collapsed to SP50, SP55 and SP102 respectively.
9. Previous review and/or resequencing of samples within the Engel et al dataset revealed the following errors: 10 haplotypes collapsed to other haplotypes, HBR003(BR57-99), HBR004(BR58-99), HBR007(BR70-99), HBR016(BR86-99), HBR045(BR152-00), HBR047(BR155-00), HBR054(BR172-01), HBR057(BR182-01), HBR062(BR192a-01) and HBR066(BR199-01): 2 samples appear to be have been switched BR108-99 and BR128-99: 1 sample, BR183-01, was heteroplasmic for a known haplotype (SP71) and a new one (HBA095): 7 samples were re sequenced as different haplotypes, BRA08-97(HBA007), BR105-99(HBA002), BRA19-98(HGA008), BRA05-98(HGA003), BRA05-97(HBA004), BRA15-97(HBA020), BRA01-97(HGA021).
10. Haplotypes and/or sequences listed in 8 and 9 above are highlighted in red in both the 'Rosetta' and the 'Engel codes' datasheets with comments listed in column D (for haplotypes) or as an attachment to the cell (for sequences).
11. Variable sites of unique haplotypes were briefly screened for potential errors such as transversions not present in other haplotypes or deletions not present in other haplotypes. Samples/haplotypes with these potential errors are bolded and comments are included in column D.
12. All haplotypes were assigned a unique GenBank and haplotype code listed in columns B and C respectively based on precedence of publication.
13. After accounting for replicate sequences and errors, the Rosetta Stone represented 223 unique haplotypes based on the consensus length of approximately 470 base pairs.

Discussion and Recommendations

The review of the 6 published datasets resulted in a final total of 223 likely haplotypes (some with potential sequencing errors), many of which are referred to by different internal laboratory codes and represented by multiple submission on GenBank (up to 16 in some popsets). Although these independent GenBank submissions were well intended and, in some cases, mandated by data archiving policies of journals, they are unwieldy.

The Rosetta Stone developed here is a first step in validating and cross-referencing haplotypes for southern hemisphere humpback whales but is known to be incomplete. Additional review will be necessary to reconcile the haplotypes in Supplementary Material of peer-reviewed publications, in manuscripts now in review or recently published and in the unpublished popsets available on GenBank. More importantly, the review did not attempt to extract or integrate information on the regional frequencies of haplotypes. This information is not included in the popsets available on GenBank and typically must be extracted from tables included in the body of the publication or as Supplementary Material, or by direct requests to authors. The potential to replicate published analyses or to conduct meta-analyses will depend on the integration of this frequency information and, if available, other metadata (e.g., year of sampling, location etc) for several thousand samples represented in the current publications.

From this initial effort, we can make the following recommendations:

- 1) The SC should agree in principle, to adopt a standardised nomenclature and consensus length for mtDNA haplotypes of humpback whales in the Southern Hemisphere and to encourage authors interested in contributing to future work of the SC to follow this

- standard. The option proposed here (Table 2) is to nominate the GenBank accession number and laboratory code for the first publication of a unique haplotypes represented by approximately 470 base pairs, e.g., following the precedence of publication. New haplotypes can be identified after first excluding a match to haplotypes in the existing dataset represented in the Rosetta Stone.
- 2) The SC should encourage authors to include information on the frequencies of haplotypes, and other metadata, in a format that could be integrated with the standardise nomenclature.
 - 3) The IWC should consider supporting a searchable, web-based ‘DNA register’ of mtDNA haplotypes for southern hemisphere humpback whales and for other species or populations of interest to future management decisions, e.g., southern hemisphere blue whales, southern right whales, North Pacific minke whales.

In making these recommendations, we recognize the need for additional computational development to improve standardisation and access. This would include an option for the automated matching of haplotypes to the existing DNA register and for selectively downloading information on regional frequencies of haplotypes. Several models exist for such searchable databases, including OBIS-SEAMAP (Halpin et al. 2006) for biogeographic datasets and www.DNASurveillance (Ross et al. 2003) for phylogenetic placement of mtDNA haplotypes. Some of the desirable attributes of an ‘intelligent’ database for integrating genetic and spatial records is available in the *Wildbook* program developed for the SPLASH project on North Pacific humpback whales, <http://www.splashcatalog.org/mmuwildbook/>.

References

- Amaral, A.R., J. Loo, H. Jaris, C. Olavarria, D. Thiele, P. Ensor, A. Aguayo and H.C. Rosenbaum. 2016. Population genetic structure among feeding aggregations of humpback whales in the Southern Ocean. *Marine Biology* 163:1-13.
- Carvalho, I., J. Loo, T. Collins, J. Barendse, C. Pomilla, M.S. Leslie, S. Ngouesso, P.B. Best and H.C. Rosenbaum. 2014. Does temporal and spatial segregation explain the complex population structure of humpback whales on the coast of West Africa? *Marine Biology* 161:805-819.
- Cypriano-Souza, A.L., M.H. Engel, S. Caballero, C. Olavarría, L. Flórez-González, J. Capella, D. Steel, A. Sremba, A. Aguayo, D. Thiele, C.S. Baker and S.L. Bonatto. In review. Genetic differentiation between humpback whales (*Megaptera novaeangliae*) from Atlantic and Pacific breeding grounds of South America.
- Dick DM, Walbridge S, Wright DJ, et al. (2014) geneGIS: Geoanalytical Tools and Arc Marine Customization for Individual-Based Genetic Records. *Transactions in GIS* 18, 324-350.
- Engel, M., N. Fagundes, H. Rosenbaum, M. Leslie, P. Ott, R. Schmitt, E. Secchi, L. Dalla Rosa and S. Bonatto. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antarctic feeding areas. *Conservation Genetics* 9:1253-1262.
- Félix, F., S. Caballero and C. Olavarría. 2012. Genetic diversity and population structure of humpback whales (*Megaptera novaeangliae*) from Ecuador based on mitochondrial DNA analyses. *Journal of Cetacean Research and Management* 12:71-77.
- Halpin, P.N., A.J. Read, B.D. Best, K.D. Hyrenbach, E. Fujioka, M.S. Coyne, L.B. Crowder, S.A. Freeman and C. Spoorri. 2006. OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles. *Marine Ecology-Progress Series* 316:239-246.
- IWC. 2015. Annex H: Report of the Sub-Committee on Other Southern Hemisphere Whale Stocks. *Journal of Cetacean Research and Management* 16 (Suppl.).
- IWC. 2016. Annex H: Report of the Sub-Committee on Other Southern Hemisphere Whale Stocks. *Journal of Cetacean Research and Management* 17 (Suppl.).

- Lang, A. (convener) et al. 2016. Standardizing nomenclature for mitochondrial DNA control region haplotypes of blue whales in the Southern Hemisphere and North Pacific. SC/66b/SH, report to the Scientific Committee of the International Whaling Commission
- Olavarria, C., C.S. Baker, C. Garrigue, M. Poole, N. Hauser, S. Caballero, L. Flórez-González, M. Brasseur, J. Bannister, J. Capella, P.J. Clapham, R. Dodemont, M. Donoghue, C. Jenner, M.N. Jenner, D. Moro, M. Oremus, D.A. Paton and K. Russell. 2007. Population structure of humpback whales throughout the South Pacific and the origins of the eastern Polynesian breeding grounds. *Marine Ecology - Progress Series* 330:257-268.
- Rosenbaum, H.C., C. Pomilla, M. Mendez, M.S. Leslie, P. B.Best, K.P. Findlay, G. Minton, P.J. Ersts, T. Collins, M.H. Engel, S. L.Bonatto, D.P.G.H. Kotze, M. Meyer, J. Barendse, M. Thornton, Y. Razafindrakoto, S. Ngouesso, M. Vely and J. Kiszka. 2009. Population Structure of Humpback Whales from Their Breeding Grounds in the South Atlantic and Indian Oceans. *PLoS ONE*.
- Rosenbaum, H.C., F.L Kershaw, M. Mendez, C. Pomilla, M.S. Leslie, K.P. Findlay, P.B. Best, T. Collins, M. Vely, M. Engel, R. Baldwin, G. Minton, M. Meyer, L. Flórez-González, M.M. Poole, N. Hauser, C. Garrigue, M. Brasseur, J. Bannister, M. Anderson, C. Olavarria, and C.S. Baker. Submitted. First circumpolar assessment of humpback whale mitochondrial genetic variation at multiple scales and implications for management.
- Ross, H.A., G.M. Lento, M.L. Dalebout, M. Goode, P. McLaren, A.G. Rodrigo, S. Lavery and C.S. Baker. 2003. DNA surveillance: web-based molecular identification of whales, dolphins and porpoises. *Journal of Heredity* 94:111-114.
- Schmitt, N.T., M.C. Double, S.N. Jarman, N. Gales, J.R. Marthick, A.M. Polanowski, C. S. Baker, D. Steel, K.C.S. Jenner and M. Jenner. 2014. Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. *Marine Mammal Science* 30:221-241.
- Valsecchi, E., P.J. Corkeron, P. Galli, W. Sherwin and G. Bertorelle. 2009. Genetic evidence for sex-specific migratory behaviour in western South Pacific humpback whales. *Marine Ecology Progress Series* 398:275.

Table 1. Summary of population datasets ('popsets') of mtDNA control region haplotypes for southern hemisphere humpback whales found on GenBank as of March 2016. Number of sequences (#seq) includes multiple submissions of haplotypes for most popsets.

Author	Year	popset link	#seq
Olavarria et al	2007	http://www.ncbi.nlm.nih.gov/popset/110798330	115
Engel et al	2008	http://www.ncbi.nlm.nih.gov/popset/168988390	253
Rosenbaum et al	2009	http://www.ncbi.nlm.nih.gov/popset/262341380	162
Valsecchi et al	2010	http://www.ncbi.nlm.nih.gov/popset/290783655	31
Felix et al	2011	http://www.ncbi.nlm.nih.gov/popset/311221575	8
Carvalho et al	2014	http://www.ncbi.nlm.nih.gov/popset/574585081	2
Other popsets available on GenBank			
Denkinger et al 1	n.a.	http://www.ncbi.nlm.nih.gov/popset/664648767	57
Denkinger et al 2	n.a.	http://www.ncbi.nlm.nih.gov/popset/528748007	18
Rojas et al	n.a.	http://www.ncbi.nlm.nih.gov/popset/974030581	25

Table 2: A "Rosetta stone" to reconcile duplicate GenBank submissions and laboratory codes for mtDNA control region haplotypes of southern hemisphere humpback whales.

#	Unique GenBank code	Unique code		Olavarria et al	Rosenbaum et al	Engel et al	Valsecchi et al	Felix et al	Carvalho et al								
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	DQ768307.1	SP1		DQ768307.1	SP1	GQ913692.1	HBA002	GQ913812.1	HGA008	AY330059.1	MN16-18						
4	DQ768310.1	SP10		DQ768310.1	SP10	GQ913716.1	HBA026			AY329901.2	BR108-99						
5	DQ768327.1	SP100		DQ768327.1	SP100					AY329967.1	MN18-19						
6	DQ768416.1	SP101		DQ768416.1	SP101												
7	DQ768330.1	SP102		DQ768330.1	SP102	GQ913732.1	HBA042					GQ872089.1	H15				
9	DQ768362.1	SP104		DQ768362.1	SP104												
10	DQ768390.1	SP105		DQ768390.1	SP105												
11	DQ768368.1	SP106		DQ768368.1	SP106												
12	DQ768355.1	SP107		DQ768355.1	SP107	GQ913739.1	HBA049	GQ913832.1	HGA029	AY329959.2	BR146-99						
13	DQ768352.1	SP108		DQ768352.1	SP108	GQ913700.1	HBA010			AY330022.2	BRA33-98	GQ872107.1	H37				
14	DQ768392.1	SP109		DQ768392.1	SP109												
15	DQ768312.1	SP11		DQ768312.1	SP11							GQ872099.1	H29				
16	DQ768394.1	SP110		DQ768394.1	SP110	GQ913740.1	HBA050										
17	DQ768309.1	SP111		DQ768309.1	SP111												
18	DQ768354.1	SP112		DQ768354.1	SP112												
19	DQ768404.1	SP113		DQ768404.1	SP113	GQ913808.1	HGA004			AY329971.2	BR113-99						
20	DQ768356.1	SP114		DQ768356.1	SP114							GQ872111.1	H41				
21	DQ768406.1	SP115		DQ768406.1	SP115												
22	DQ768341.1	SP12		DQ768341.1	SP12												
23	DQ768324.1	SP13		DQ768324.1	SP13	GQ913744.1	HBA058	GQ913696.1	HBA006	AY329880.2	BR175-01	GQ872097.1	H27				
24	DQ768318.1	SP14		DQ768318.1	SP14	GQ913761.1	HBA083	GQ913824.1	HGA020	AY329882.2	BR173-01						
25	DQ768334.1	SP15		DQ768334.1	SP15	GQ913761.1	HBA083	GQ913824.1	HGA020	AY329882.2	BR173-01						
26	DQ768410.1	SP16		DQ768410.1	SP16												
27	DQ768338.1	SP17		DQ768338.1	SP17												
28	DQ768367.1	SP18		DQ768367.1	SP18												
29	DQ768371.1	SP19		DQ768371.1	SP19	GQ913836.1	HGA038					GQ872095.1	H24				
30	DQ768308.1	SP2		DQ768308.1	SP2												
31	DQ768391.1	SP20		DQ768391.1	SP20												
32	DQ768401.1	SP21		DQ768401.1	SP21												
33	DQ768369.1	SP22		DQ768369.1	SP22												
34	DQ768387.1	SP23		DQ768387.1	SP23	GQ913711.1	HBA021										
35	DQ768380.1	SP24		DQ768380.1	SP24	GQ913780.1	HBA111										
36	DQ768386.1	SP25		DQ768386.1	SP25	GQ913712.1	HBA022			AY329885.2	BR149-99						
37	DQ768340.1	SP26		DQ768340.1	SP26	GQ913738.1	HBA048										
38	DQ768313.1	SP27		DQ768313.1	SP27	GQ913707.1	HBA017			AY330026.2	BRA27-98	GQ872094.1	H23				
39	DQ768337.1	SP28		DQ768337.1	SP28												
40	DQ768342.1	SP29		DQ768342.1	SP29	GQ913726.1	HBA036					GQ872109.1	H39				
41	DQ768314.1	SP3		DQ768314.1	SP3	GQ913774.1	HBA100					GQ872106.1	H36				
42	DQ768388.1	SP30		DQ768388.1	SP30												
43	DQ768384.1	SP31		DQ768384.1	SP31												
44	DQ768414.1	SP32		DQ768414.1	SP32					AY330083.1	MN57-17						
45	DQ768361.1	SP33		DQ768361.1	SP33	GQ913792.1	HBR035			AY329907.2	BR132-99						
46	DQ768396.1	SP34		DQ768396.1	SP34												
47	DQ768405.1	SP35		DQ768405.1	SP35	GQ913714.1	HBA024	GQ913819.1	HGA015	AY329889.2	BR181-01						
48	DQ768326.1	SP36		DQ768326.1	SP36	GQ913776.1	HBA102										
49	DQ768375.1	SP37		DQ768375.1	SP37							GQ872104.1	H34				
50	DQ768333.1	SP38		DQ768333.1	SP38	GQ913709.1	HBA019	GQ913733.1	HBA043	AY330010.2	BRA29-99						

196	GQ913772.1	HBA098		GQ913772.1	HBA098				
145	GQ913773.1	HBA099		GQ913773.1	HBA099		AY329855.2	BR87-99	
146	GQ913775.1	HBA101		GQ913775.1	HBA101		AY329881.2	BR166-01	
216	GQ913777.1	HBA103		GQ913777.1	HBA103	GQ913782.1 HBA113			
122	GQ913781.1	HBA112		GQ913781.1	HBA112		AY329985.2	BRA16-98	
147	GQ913783.1	HBR002		GQ913783.1	HBR002		AY330025.2	BRA03-98	
174	GQ913788.1	HBR014	SP71 without transversion	GQ913788.1	HBR014		AY329859.2	BR83-99	
148	GQ913790.1	HBR022		GQ913790.1	HBR022		AY329854.2	BR93-99	
149	GQ913791.1	HBR028		GQ913791.1	HBR028		AY329893.2	BR107-99	
150	GQ913793.1	HBR039		GQ913793.1	HBR039		AY329931.2	BR143-99	
192	GQ913802.1	HEZ001	HBA038 without transversion	GQ913802.1	HEZ001				
195	GQ913803.1	HEZ002		GQ913803.1	HEZ002				
177	GQ913804.1	HEZ003		GQ913804.1	HEZ003				
199	GQ913805.1	HGA001		GQ913805.1	HGA001				
254	GQ913806.1	HGA002		GQ913806.1	HGA002				
151	GQ913809.1	HGA005		GQ913809.1	HGA005		AY329909.2	BR106-99	
129	GQ913811.1	HGA007		GQ913811.1	HGA007		AY329891.2	BR125-99	
203	GQ913813.1	HGA009		GQ913813.1	HGA009				
205	GQ913814.1	HGA010		GQ913814.1	HGA010				
152	GQ913815.1	HGA011		GQ913815.1	HGA011		AY329900.2	BR103-99	
206	GQ913816.1	HGA012		GQ913816.1	HGA012				
207	GQ913817.1	HGA013		GQ913817.1	HGA013				
153	GQ913818.1	HGA014		GQ913818.1	HGA014		AY329899.2	BR202-01	
209	GQ913820.1	HGA016		GQ913820.1	HGA016				
136	GQ913821.1	HGA017		GQ913821.1	HGA017	GQ913840.1 HGA042	AY329888.2	BR99-99	
210	GQ913822.1	HGA018		GQ913822.1	HGA018				
211	GQ913823.1	HGA019		GQ913823.1	HGA019				
213	GQ913826.1	HGA022		GQ913826.1	HGA022		AY329984.1	BRA01-98	
214	GQ913827.1	HGA023		GQ913827.1	HGA023				
215	GQ913828.1	HGA024		GQ913828.1	HGA024				
217	GQ913830.1	HGA026		GQ913830.1	HGA026				
218	GQ913831.1	HGA027	HBA063 without transversion	GQ913831.1	HGA027				
219	GQ913833.1	HGA034	SP108 without transversion	GQ913833.1	HGA034				
154	GQ913834.1	HGA036		GQ913834.1	HGA036		AY329927.2	BR150-99	
137	GQ913835.1	HGA037		GQ913835.1	HGA037		AY329898.2	BR147-99	GQ872105.1 H35
251	GQ913837.1	HGA039	HGA007 without transversion	GQ913837.1	HGA039				
221	GQ913838.1	HGA040		GQ913838.1	HGA040				
189	GQ913841.1	HMY001		GQ913841.1	HMY001				
198	GQ913843.1	HMZ002		GQ913843.1	HMZ002				
241	GQ913844.1	HMZ005		GQ913844.1	HMZ005				
197	GQ913845.1	HOM001		GQ913845.1	HOM001				
202	GQ913846.1	HOM002		GQ913846.1	HOM002				
212	GQ913847.1	HOM003		GQ913847.1	HOM003				
220	GQ913848.1	HOM004		GQ913848.1	HOM004				
190	GQ913849.1	HWZ001		GQ913849.1	HWZ001				
138	GQ913850.1	HWZ002		GQ913850.1	HWZ002		AY329906.2	BR190-01	
194	GQ913851.1	HWZ005		GQ913851.1	HWZ005				
155	GQ913852.1	HWZ006		GQ913852.1	HWZ006		AY329886.2	BR156-00	
124	AY329926.2	MN03-19					AY329926.2	MN03-19	
250	AY330072.1	MN08-17	deletion at site unique to Engel dataset				AY330072.1	MN08-17	
242	AY330032.1	MN11-17	deletion at site unique to Engel dataset				AY330032.1	MN11-17	
234	AY330039.2	MN26-17					AY330039.2	MN26-17	
235	AY330040.2	MN33-17	unique without transversion				AY330040.2	MN33-17	
260	AY330094.2	MN50-17	SP17 without transversion				AY330094.2	MN50-17	

246	AY330092.2	MN54-17	SP63 without transversion			AY330092.2	MN54-17				
243	AY330062.2	MN66-17	2 deletions at sites unique to Engel dataset			AY330062.2	MN66-17				
258	GQ872085.1	H08						GQ872085.1	H08		
252	GQ872096.1	H26						GQ872096.1	H26		
131	GQ872098.1	H28						GQ872098.1	H28		
244	GQ872112.1	H42						GQ872112.1	H42		
239	HQ241479.1	EC001								HQ241479.1	EC001
240	HQ241480.1	EC002								HQ241480.1	EC002
245	HQ241481.1	EC003								HQ241481.1	EC003
238	HQ241482.1	EC004								HQ241482.1	EC004
125	HQ241483.1	EC005								HQ241483.1	EC005
236	HQ241484.1	EC006								HQ241484.1	EC006
262	HQ241485.1	EC007								HQ241485.1	EC007
255	HQ241486.1	EC008	unique without transversion							HQ241486.1	EC008
8	DQ768378.1	SP103	collaps	DQ768378.1	SP103						
65	DQ768363.1	SP51	collaps	DQ768363.1	SP51						
70	DQ768358.1	SP56	collaps	DQ768358.1	SP56						
222	GQ913784.1	HBR003	reseq'd as HBA002		SP1	GQ913784.1	HBR003	AY329961.2	BR57-99		
223	GQ913785.1	HBR004	reseq'd as HBA112			GQ913785.1	HBR004	AY329962.2	BR58-99		
172	GQ913787.1	HBR007	reseq'd as HBR035		SP33	GQ913787.1	HBR007	AY330095.2	BR70-99		
175	GQ913789.1	HBR016	reseq'd as HBA002		SP1	GQ913789.1	HBR016	AY329970.2	BR86-89		
179	GQ913794.1	HBR045	reseq'd to HBA065			GQ913794.1	HBR045	AY329914.2	BR152-00		
224	GQ913795.1	HBR047	reseq'd as HGA037			GQ913795.1	HBR047	AY329963.2	BR155-00		
225	GQ913797.1	HBR054	reseq'd to HBA049		SP107	GQ913797.1	HBR054	AY329960.2	BR172-01		
184	GQ913798.1	HBR057	reseq'd to HGA017		SP10	GQ913798.1	HBR057	AY329879.2	BR182-01		
226	GQ913799.1	HBR062	reseq'd to HBA012			GQ913799.1	HBR062	AY329964.2	BR192a-01		
187	GQ913801.1	HBR066	reseq'd to HBA011		SP94	GQ913801.1	HBR066	AY329933.2	BR199-01		