Y Chromosome Analysis of Dingoes and Southeast Asian Village Dogs Suggests a Neolithic Continental Expansion from Southeast Asia Followed by Multiple Austronesian Dispersals

Benjamin N. Sacks,*[,1,2] Sarah K. Brown,[1] Danielle Stephens,[‡,3] Niels C. Pedersen,[4,5] Jui-Te Wu,[6] and Oliver Berry[7]

[1]Canid Diversity and Conservation Unit, Veterinary Genetics Laboratory, University of California, Davis

[2]Department of Population, Health and Reproduction, School of Veterinary Medicine, University of California, Davis

[3]Invasive Animals Cooperative Research Centre, School of Animal Biology (M092), University of Western Australia, Crawley, Western Australia, Australia

[4]Center for Companion Animal Health, School of Veterinary Medicine, University of California, Davis

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis

[6]Department of Veterinary Medicine, National Chiayi University, Chiayi City, Taiwan, Republic of China

[7]CSIRO Marine and Atmospheric Research, Centre for Environment and Life Sciences, Floreat, Western Australia, Australia

[‡]Present address: Helix Molecular Solutions, Crawley, Western Australia, Australia

***Corresponding author:** E-mail: bnsacks@ucdavis.edu.

Associate editor: Matthew Hahn

Abstract

Dogs originated more than 14,000 BP, but the location(s) where they first arose is uncertain. The earliest archeological evidence of ancient dogs was discovered in Europe and the Middle East, some 5–7 millennia before that from Southeast Asia. However, mitochondrial DNA analyses suggest that most modern dogs derive from Southeast Asia, which has fueled the controversial hypothesis that dog domestication originated in this region despite the lack of supporting archeological evidence. We propose and investigate with Y chromosomes an alternative hypothesis for the proximate origins of dogs from Southeast Asia—a massive Neolithic expansion of dogs from this region that largely replaced more primitive dogs to the west and north. Previous attempts to test matrilineal findings with independent patrilineal markers have lacked the necessary genealogical resolution and mutation rate estimates. Here, we used Y chromosome genotypes, composed of 29 single-nucleotide polymorphism (SNPs) and 5 single tandem repeats (STRs), from 338 Australian dingoes, New Guinea singing dogs, and village dogs from Island Southeast Asia, along with modern European breed dogs, to estimate the evolutionary mutation rates of Y chromosome STRs based on calibration to the independently known age of the dingo population. Dingoes exhibited a unique haplogroup characterized by a single distinguishing SNP mutation and 14 STR haplotypes. The age of the European haplogroup was estimated to be only 1.7 times older than that of the dingo population, suggesting an origin during the Neolithic rather than the Paleolithic (as predicted by the Southeast Asian origins hypothesis). We hypothesize that isolation of Neolithic dogs from wolves in Southeast Asia was a key step accelerating their phenotypic transformation, enhancing their value in trade and as cargo, and enabling them to rapidly expand and replace more primitive dogs to the West. Our findings also suggest that dingoes could have arrived in Australia directly from Taiwan, independently of later dispersals of dogs through Thailand to Island Southeast Asia.

Key words: **Austronesian expansion, dingo, dog, evolutionary mutation rate, single tandem repeat, Y chromosome.**

Introduction

Tracking Holocene movements of domesticated plants and animals with molecular genetic markers has provided important insights into prehistoric human migrations (Dobney and Larson 2006; Zeder et al. 2006; Matisoo-Smith et al. 2007; Larson et al. 2010). Dogs, the first domesticated animal, were associated with Neolithic humans on most continents, making them an especially valuable proxy for understanding Holocene human movements (Clutton-Brock 1995; Leonard et al. 2002; Savolainen et al. 2002, 2004; Morey 2010; Brown et al. 2013). However, the low resolution of mitochondrial DNA (mtDNA) markers over recent historical timescales has presented a challenge in mid- to late-Holocene phylogenetic studies. Moreover, inconsistent geographic representation of archeological faunal remains, which are needed along with DNA evidence to reconstruct late-Pleistocene and early-Holocene migrations (Zeder et al. 2006), have limited our understanding of phylogeography of the earliest dogs.

Establishing the origins of the first dogs has been a particularly elusive task. On the one hand, the earliest archeological evidence of dogs, dating more than 11,000

Article

years before present (BP), is distributed throughout Europe and the Middle East, possibly as far east as Kamchatka (e.g., Olsen 1985, Clutton-Brock 1995; Dikov 1996; Tchernov and Valla 1997; Pionnier-Capitan et al. 2011; Crockford and Kuzmin 2012). However, dog remains are unknown before 7,100 BP from Southeast Asia (Underhill 1997; Larson et al. 2012), implying a west Eurasian origin. On the other hand, worldwide analyses of mtDNA indicate that modern dog matrilines reflect a subset of those found today in Southeast Asia (Savolainen et al. 2002; Pang et al. 2009; Ardalan et al. 2011; Brown et al. 2011). This observation has led to the "Southeast Asian origins" hypothesis that the earliest dogs were derived from a now-extinct wolf from south of the Yangtzee River in China. The lack of supporting archeological evidence has been explained by the paucity of archeological sites and historically low interest in and attention to faunal remains in Asia (Savolainen et al. 2002; Pang et al. 2009).

If, however, dogs were absent from Southeast Asia until the Neolithic, the most parsimonious hypothesis reconciling archeological and mtDNA observations would be that early dogs entered Southeast Asia with pre-Neolithic peoples from the west or north, then later expanded outward 8–5 Ka swamping or replacing more primitive dog populations. Several lines of evidence support such a "Neolithic replacement" hypothesis. First, Neolithic expansion of dogs from Southeast Asia 8–5 Ka would be in line with linguistic, cultural, archeological, and human genetic evidence of a westward expansion of Neolithic humans from the Yangtzee and Yellow River basins (Diamond and Bellwood 2003), the precursor to the slightly later migrations of Austronesian-speaking farmers from this same region, which were responsible for spreading dogs to Island Southeast Asia and Oceana (Savolainen et al. 2004; Brown et al. 2011; Oskarsson et al. 2011). Archeological and ancient DNA evidence suggesting that late-Paleolithic dogs were replaced by Neolithic immigrants in regions as disparate as Japan, the Middle East, and North America also support this hypothesis (Olsen 1985; Tanabe 2006; Tchernov and Valla 1997; Brown et al. 2013).

Molecular genetic tools offer an independent means of testing whether modern western dog lineages diverged from their Southeast Asian progenitors in Paleolithic timeframes, consistent with the "Southeast Asian Origins" hypothesis (Pang et al. 2009), or in later Neolithic timeframes, in support of the "Neolithic Replacement" hypothesis. However, the maternally inherited mtDNA marker used for most phylogeographic studies of the dog mutates too slowly to distinguish these timeframes (Savolainen et al. 2002; Pang et al. 2009). For example, the mtDNA-based estimate of dingo arrival in Australia was estimated at 4.5–11 Ka (Savolainen et al. 2004), which is similar to the estimated divergence between European and Southeast Asian dogs (5.5–16.5 Ka; Pang et al. 2009) (although imprecision of these estimates was partly due to incorporation of a 2-fold range of assumed mutation rates). Thus, a more rapidly mutating, phylogenetically tractable marker on clonally inherited DNA is needed to provide greater precision.

For human genetic studies, the approach of combining rapidly mutating single tandem repeats (STRs) with slower mutating single-nucleotide polymorphisms (SNPs) on the nonrecombining region of the Y chromosome has provided a high-resolution alternative (and complement) to mtDNA that is better suited to resolving Holocene time scales (Heyer et al. 1997; Jobling and Tyler-Smith 2003). The SNPs provide a deep phylogenetic scaffold on which the rapidly mutating STRs are phylogenetically organized (i.e., into haplogroups) and the STR haplotypes, once so-organized, provide the rapid "molecular clock." Phylogeographic analysis of human Y chromosome variation has been facilitated by calibrations of mutation rates in various isolated human populations (Forster et al. 2000; Kayser et al. 2000; Zhivotovsky et al. 2004; Shi et al. 2009). These calibrated mutation rates have, in turn, enabled inferences about human movements as recently as 2,000 years. For example, Henn et al. (2008) discovered a Y chromosome SNP that unified a formerly paraphyletic Y STR haplogroup, enabling them to use phylogeographically tractable Y STR haplotypes to infer the timing and source of a southern African human migration. Although studies of dogs and other canids have used Y chromosomes to a limited extent (e.g., Bannasch et al. 2005; Sacks et al. 2008), only two have attempted to infer demographic history of dogs (Brown et al. 2011; Ding et al. 2011).

Both dog Y chromosome studies addressed phylogeography in Eurasia but reached fundamentally different conclusions about the divergence time between eastern and western dogs based on differing and poorly tested assumptions about mutation rates of the markers used in the two studies. In particular, one study using SNPs (substitutions from 14,000 base pairs [bp] of 160 dog Y chromosomes) found European and Southeast Asian dogs to reflect primarily different haplogroups but lacked resolution to age the diagnostic European mutation (Ding et al. 2011). The other study used STRs to find that European and Southeast Asian dogs clustered together in a "Southeast Asian clade," which was highly differentiated from a "Middle Eastern clade," but it lacked the necessary deeper phylogenetic scaffolding afforded by SNPs and knowledge of the STR mutation rate to determine whether this clustering reflected Neolithic-age connections or post-Victorian exchange of European and Southeast Asian stock in the creation of modern breeds (Brown et al. 2011).

Knowing mutation rates for these markers, especially the STRs, is necessary to the estimation of divergence times. Mutation rates have been estimated for a number of autosomal dog STRs based on direct counts of repeat changes accumulating in multigenerational pedigrees or father–son pairs (Irion et al. 2003; Parra et al. 2010), which generally agree with similarly estimated mutation rates observed in human autosomal STRs lineages (Weber and Wong 1993). Moreover, in humans, Y chromosome STRs have similar mutation rates to their autosomal counterparts (Heyer et al. 1997). However, the rates of mutation as inferred from phylogenetic data (i.e., "evolutionary mutation rates" [EMRs]) can be an order of magnitude lower than those observed directly from pedigrees (Forster et al. 2000; Zhivotovsky et al. 2004). For this reason, it is important to calibrate EMRs to

populations of known age (Forster et al. 2000; Zhivotovsky et al. 2004; Shi et al. 2009).

The objectives of this study were to combine the dog Y chromosome SNPs and STRs to resolve phylogenetic relationships among patrilines, to estimate EMRs of the STRs through calibration of a large sample of Australian dingoes (*Canis lupus dingo*) in conjunction with Bali and other Southeast Asian village dogs, and to revisit hypotheses about the timing of divergence between modern European and Southeast Asian dogs. Secondarily, we sought to explore the implications of a calibrated Y-chromosome "molecular clock" for the chronology of the Austronesian expansion. Dingoes are ideal for calibration of EMRs as they reflect a known 5,000–3,500 years of isolation and evolution from as few as 2–4 founders based on archeological and biogeographic evidence and consistent maternal genetic evidence (Corbett 1995; Savolainen et al. 2004; Oskarsson et al. 2011; Ardalan et al. 2012; Fillios et al. 2012; Larson et al. 2012). Although domestic dogs introduced to Australia in the last two centuries are known to interbreed with dingoes, especially in the Southeast of the continent, we used dingoes from remote regions of Australia where such interbreeding was rare and, in particular, used individuals that had been screened using autosomal nuclear genetic criteria for "purity" before this study (Elledge et al. 2008; Stephens 2011), which we also confirmed in this study through identification of indicator mutations in both the mtDNA and Y chromosome (Savolainen et al. 2004; Ardalan et al. 2012).

Results

Dingo mtDNA

We observed 15 D loop haplotypes in 94 dingoes (89 males and 5 females), of which nine were novel (not previously described; table 1). Ignoring indels, this was equivalent to 10 distinct 582-bp haplotypes, of which four were novel. Previously described dingo haplotypes (Savolainen et al. 2004) were found in 86 individuals, including 36 exhibiting the presumed founding haplotype (A29). The four novel haplotypes were found in 1–2 individuals and differed by 1–2 substitutions from A29, which was found in 36 individuals. Otherwise, two individuals contained haplotypes that were previously found in nondingo dogs, including A9, which probably evolved independently in dingoes (Savolainen et al. 2004). Thus, only one of the 94 dingoes used in this study carried a haplotype likely to have originated from a domestic dog (A17). Adding the haplotypes in this study to previously published ones yielded a total of 336 dingo sequences, which continue to exhibit a star-like genealogical structure consistent with the expansion of dingoes from as few as a single female founder (fig. 1).

Table 1. Mitochondrial D Loop Haplotypes Observed in 94 Dingoes.

Haplotype[a]	Excluding Indels	*N*	Inferred Origin	GenBank No.
A9	A9	1	Dingo derived[b]	AY660645
A17	A17	1	Introgression	AF531669
A29	A29	27	Founder	AY660633
din26[a]	A29	1	Dingo derived	KC106734
din28[a]	A29	2	Dingo derived	KC106736
A179	A29	6	Dingo derived	AY660651
din24[a]	A200	2	Dingo derived	KC106733
A200	A200	16	Dingo derived	AY660635
A201	A201	1	Dingo derived	AY660636
A203	A203	31	Dingo derived	AY660638
din29[a]	A203	1	Dingo derived	KC106737
din22[a]	din22[a]	2	Dingo derived	KC106731
din23[a]	din23[a]	1	Dingo derived	KC106732
din27[a]	din27[a]	1	Dingo derived	KC106735
din30[a]	din30[a]	1	Dingo derived	KC106738

[a]Haplotypes din22–din30 were first described in this study.

[b]Although haplotype A9 is widespread in dogs outside Australia, its occurrence in dingoes likely reflects an independent derivation in situ (Savolainen et al. 2004).

Y Chromosome Haplotypes

We observed 85 Y chromosome SNP–STR haplotypes in 338 dogs, dingoes, and New Guinea singing dogs (NGSDs; supplementary table S1, Supplementary Material online). In the 112 dingoes and NGSDs typed in this study (including 5 male Fraser Island dingoes from Brown et al. 2011), we observed 20 STR haplotypes clustered in three haplogroups (i.e., corresponding to distinct SNP haplotypes). Most dingoes and all NGSDs carried the H60 SNP mutation apparently unique to these populations (Ardalan et al. 2012). Correspondingly, most of the dingo/NGSD STR haplotypes were novel and all of them clustered in the "Southeast Asian clade" (sensu Brown et al. 2011) (supplementary tables S1 and S2, Supplementary Material online).

Altogether, the 338 Y chromosomes examined in this study included 10 of the 18 SNP haplotypes occurring in the "Southeast Asian clade," representing all but one of those previously observed in Southeast Asian dogs, dingoes, or NGSDs (Ding et al. 2011; Ardalan et al. 2012). Except for H15, which had been reported in 3 Chinese dogs, the other potential haplotypes (H11, H16, H19, H21, H22, H24, and H31) were absent from Australasia in this study and previous ones (Ding et al. 2011; Ardalan et al. 2012).

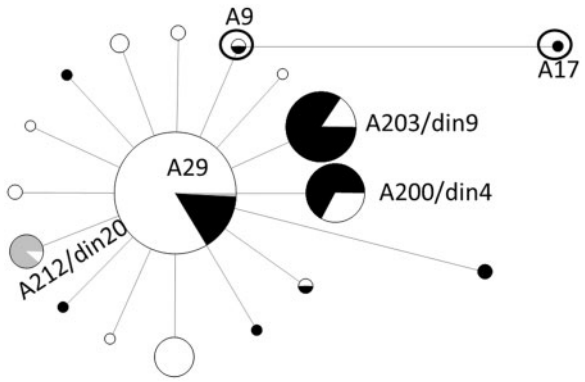

FIG. 1. Mitochondrial D loop haplotypes (582 bp, ignoring indels) showing all 336 dingoes from our study (black, $n = 94$), Brown et al. (2011, gray, $n = 10$), and previous ones (white, $n = 232$; Savolainen et al. 2004; Elledge et al. 2008; Oskarsson et al. 2011).

Y Chromosome Networks

Establishing the relationships among the 29 SNP markers enabled us to clarify the phylogenetic relationships among the more rapidly mutating STRs and to link these results to those from the previous analysis of STRs and a subset of these SNPs (Brown et al. 2011). For example, two cases were evident where STR haplotypes were identical in state but not by descent (supplementary fig. S1, Supplementary Material online). Bali dogs and dingoes were previously found to share STR haplotype "0i" (Brown et al. 2011), but our use of 29 SNPs here indicated that these identical STR haplotypes clustered in different haplogroups corresponding, respectively, to SNP types H3 and H60. Similarly, the STR haplotype n4 recorded from both dingoes and breed dogs corresponded to SNP types H60 in dingoes but H1 in a breed dog. Conversely, two other haplotypes sharing STR profiles but differing by a SNP mutation more likely reflected recent SNP substitutions than STR homoplasy (0n-H3, 0n-H12; 6r-H3, and 6r-H4). Lastly, STR haplotype 0m corresponded to adjacent but internal positions on the network (H6 in Bali and H3 in Iran), suggesting either an uncharacteristic degree of STR conservatism or, more likely, an accumulation of STR back-mutations.

The SNP–STR haplotype network composed of all 338 Y chromosomes indicated substantial phylogenetic structure, most notably, distinctive clustering of dingo haplotypes and, to some extent, Bali dog haplotypes (fig. 2A). Also importantly, NGSDs shared two haplotypes with dingoes, suggesting these were >2,000 years old—the latest time by which convincing archeological remains of dogs (i.e., other than teeth, which predate this but could have been traded) are known on New Guinea (Bulmer 2001)—and possibly included the original dingo (and NGSD) founder haplotype. A number of nondingo haplotypes also were shared across populations.

It is noteworthy that 9 of 13 (69%) of the remaining shared haplotypes were observed in breed dogs and, in particular, that eight shared haplotypes occurred in haplogroup H1, which was the predominant haplogroup among breed dogs in this study (76% of breed dog haplotypes) and among European dogs (also breed dogs) in a previous study (Ding et al. 2011). In fact, all the H1 haplotypes in the village dogs that had also been sampled in breed dogs were from European and American, but not Asian, breeds (Brown et al. 2011; Pedersen et al. 2012). Specific haplotype associations with breeds included 6q in English Setter ($n = 16$), Brittany ($n = 1$), Italian Greyhound ($n = 2$), Brussels Griffon ($n = 1$), and West Highland White Terrier ($n = 1$); 8d in Italian Greyhound ($n = 42$); 6p in English and Red Setters ($n = 13$), German Shorthairs and Wirehairs ($n = 27$), and Standard Poodle ($n = 91$); 6zi in Red Setter ($n = 6$), Golden Retriever ($n = 3$), and Schnauzer ($n = 1$); 1c in Labrador Retriever ($n = 1$); and 7d in Shetland Sheepdog ($n = 1$).

The proportion of haplotypes shared across populations and the proportion of haplotypes in haplogroup H1 also were highly correlated among nonbreed populations ($r = 0.94$), further suggesting that translocation of western breed dogs was the primary cause of haplotype sharing across populations (fig. 2C). Proportions of haplotypes shared across populations or in the H1 haplogroup were highest in Taiwan (75%, 75%, respectively), Brunei (67%, 55%), the Philippines (40%, 50%), and Thailand (45%, 27%), with much lower proportions of these quantities observed in Bali dogs (13%, 6%) and dingoes (10%, 10%). The lower proportions on Bali are consistent with the ban on importation of dogs there established in 1926 (Irion et al. 2005). Likewise, although dog–dingo interbreeding is known to be common in Southeastern Australia (Elledge et al. 2008; Stephens 2011), our deliberate use of dingoes from remote areas and our genetic screening for purity before inclusion in this study was apparently effective in minimizing non-native haplotypes in our dingo sample as well. The phylogenetic distinctions among populations were evident when breed dogs were excluded from the network as was the exceptional nature of haplogroup H1 with respect to haplotype sharing among these populations (fig. 2B).

Mutation Rate Estimates

Identification of ancestral nodes of haplotype clusters corresponding to dingoes (H60 haplogroup) and Bali dogs (a portion of H3) was uncertain due to homoplasy among STR types. This was especially problematic in the Bali sample because of the unknown degree of isolation of that population and the close association of haplotypes with those of other Southeast Asian populations, which suggested the possibility that the cluster reflected multiple founding events. Therefore, we calculated estimates that corresponded to a range of plausible ancestral nodes in both populations and, additionally, assumed both one and two founders for the H3 haplotype cluster of the Bali sample.

The estimates were relatively consistent regardless of which node was assumed ancestral but, not surprisingly, differed somewhat for the Bali population depending on whether one or two founders were assumed (tables 2 and 3). The Bali estimates assuming a single founder were nearly identical to the dingo-based estimates. Although average squared difference (ASD; Goldstein et al. 1995) estimates were higher on average than ρ estimates (Forster et al. 2000), these differences were relatively small compared with the differences due to assumed generation time and time since founding of the respective populations. For example, the average dingo ρ-based estimated EMRs for the 5-locus haplotype (1.99, standard error [SE] = 0.69) corresponded to 4.0×10^{-4} (SE = 1.4×10^{-4}) per year, assuming a 5,000-year time to most recent common ancestor (TMRCA), or 5.7×10^{-4} (SE = 2.0×10^{-4}) per year assuming a 3,500-year TMRCA. For comparison, we also estimated the ρ age of the H1 clade (fig. 2A) to get an idea of the relative age of this predominantly European haplogroup. The 5-locus haplotype estimate, $\rho = 3.34$ (SE = 1.00), was approximately 1.7 times greater than that for the dingo H60 haplogroup, corresponding to time estimates ranging 5,874 years (SE = 1,756 years), assuming a dingo founding time of 3,500 years, or 8,392 (SE = 2,513) years, assuming a dingo founding time of 5,000 years.

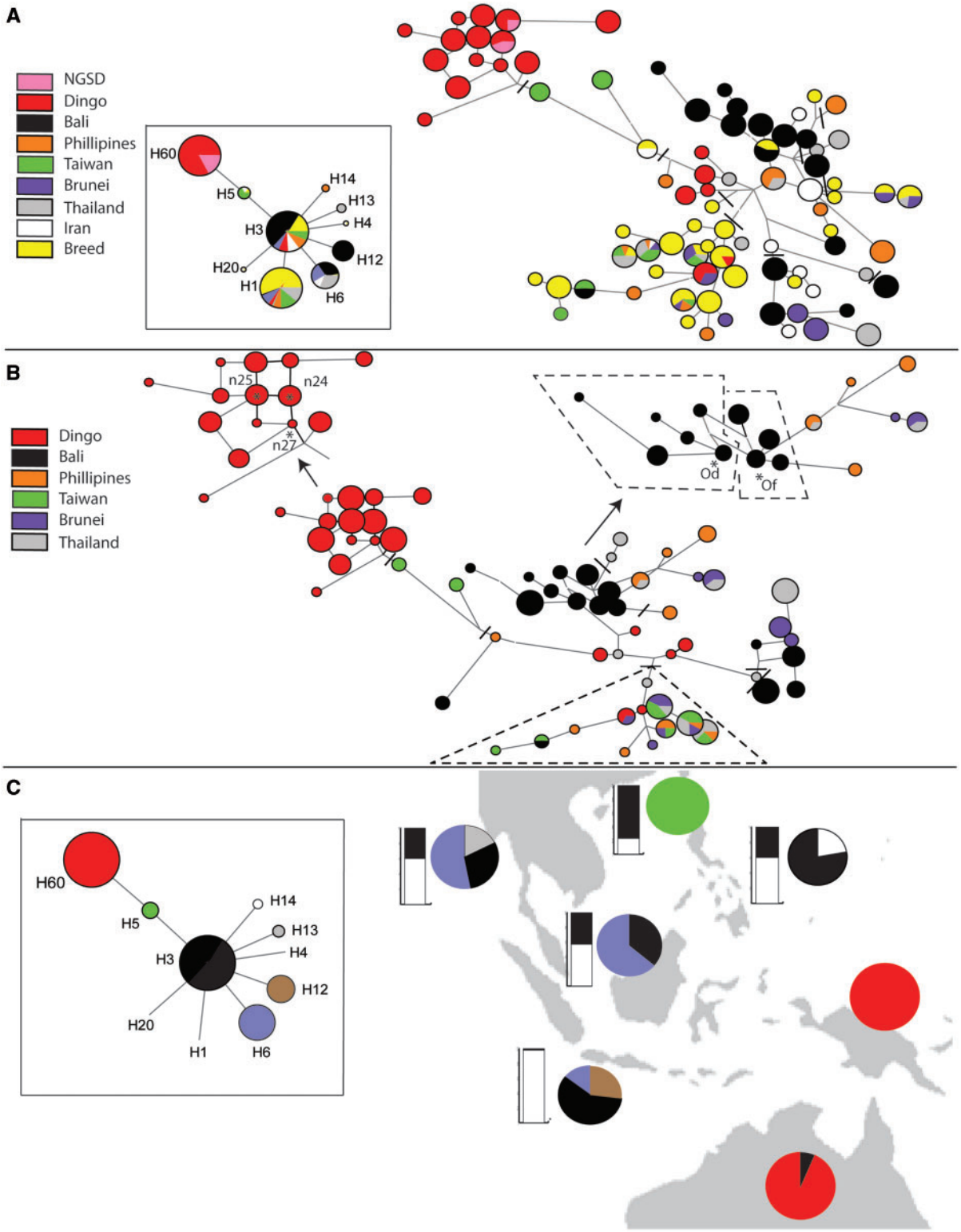

FIG. 2. MJ networks of Y chromosome haplotypes composed of 29 SNPs and 5 STRs (*A*) among 85 haplotypes in 338 breed dogs, village dogs, and dingoes, (*B*) 74 haplotypes in 237 Southeast Asian village dogs and Australian dingoes, and (*C*) haplogroup composition of geographically mapped samples. Network circle size corresponds approximately to sample size, and connection lengths are proportional to numbers of mutations. Perpendicular hash marks indicate substitutions (SNPs) distinguishing haplogroups corresponding to SNP haplotypes. Insets show (*A*) network of haplotypes based on SNPs only and (*B*) portions of network corresponding to haplogroups (H60 for dingoes, H3 for Bali dogs) used to estimate mutational accumulation (ρ) in dingo and bali populations, with STR haplotype names indicated for putative ancestral nodes (*). Dashed ellipses indicate subclades used in the Bali analysis, and the dashed polygon indicates haplogroup H1, which was excluded from the Bayesian analysis of splitting times among Australasian populations. (*C*) Vertical bars indicate proportions of dog samples composed of putative indigenous haplotypes (white) versus putative western haplotypes (H1) (black); pie charts reflect composition of indigenous SNP haplotypes with respect to the color-coded network to the left.

Table 2. Estimates of Per-Year EMRs Based on Genealogical Mutational Accumulation (ρ) and ASD in Y Chromosome STR Haplotypes within the H60 Haplogroup of Dingoes ($n = 85$) Corresponding to Several Putative Ancestral and Derived Nodes.

Ancestral Node (Reason Presumed)	Haplotype (5-loc) Est. ρ (SE)	ρ, Per-Locus EMR[a] ($\times 10^{-4}$)		Est. ASD, EMR (SE)	ASD, Per-Locus EMR ($\times 10^{-4}$)	
		3,500-Year TMRCA, EMR (SE)	5,000-Year TMRCA, EMR (SE)		3,500-Year TMRCA, EMR (SE)	5,000-Year TMRCA, EMR (SE)
n24 (shared, freq)	1.8 (0.5)	0.7 (0.5)	1.0 (0.7)	0.6 (0.7)	1.3 (1.3)	2.0 (1.9)
n24 (shared, freq)	1.5 (0.5)[b]	0.6 (0.5)	0.9 (0.7)	0.5 (0.5)	1.0 (1.0)	1.4 (1.4)
n27 (basal)	2.3 (0.8)	1.0 (0.8)	1.3 (1.1)	0.8 (0.9)	1.5 (1.7)	2.2 (2.5)
n27 (basal)	2.3 (0.9)[b]	1.0 (0.9)	1.3 (1.2)	0.8 (0.8)	1.6 (1.7)	2.2 (2.4)
n25 (central)	1.9 (0.6)	0.8 (0.6)	1.1 (0.8)	0.7 (0.7)	1.5 (1.5)	2.1 (2.1)
Average		0.8 (0.7)	1.1 (0.9)		1.3 (1.4)	1.9 (2.0)

[a]The ρ-based EMR estimate is based on the haplotype estimates divided by 5 loci; the associated haplotype SE was converted to variance (squared), divided by 5 loci to produce the average per-locus variance estimate, which was then converted to the average per-locus SE (square root of variance) estimate.
[b]Estimates excluding haplotypes n18 and n29, which appeared basal to A24 and A27.

Table 3. Estimates of Per-Year EMRs Based on Genealogical Mutational Accumulation (ρ) and ASD in Y Chromosome STR Haplotypes within the H3 Haplogroup of Bali Dogs ($n = 37$) Corresponding to Several Putative Ancestor-Descendent Relationships among Haplotypes and Assuming One or Two Founding Haplotypes.

Ancestral Node(s) (Reason Presumed)	Derived Nodes Presumed	Haplotype (5-loc) Est., ρ (SE)	ρ, Per-Locus EMR[a] ($\times 10^{-4}$)		Est. ASD, Avg. (SE)	ASD, Per-Locus EMR ($\times 10^{-4}$)	
			3,000-Year TMRCA, Avg. (SE)	4,500-Year TMRCA, Avg. (SE)		3,000-Year TMRCA, Avg. (SE)	4,500-Year TMRCA, Avg. (SE)
Single founder							
0f (basal, shared)	0a-0e,0g-0j	1.97 (0.77)	0.9 (0.8)	1.3 (1.2)	0.52 (0.42)	1.2 (1.0)	1.7 (1.4)
0d (basal)	0a-0c,0e-0j	2.09 (0.77)	1.0 (0.8)	1.4 (1.2)	0.54 (0.43)	1.2 (1.0)	1.8 (1.4)
	Average		1.0 (0.8)	1.4 (1.2)		1.2 (1.0)	1.8 (1.4)
Two founders							
0f,0d	0f: 0b, 0c, 0h, 0j	1.16 (0.42)	0.5 (0.4)	0.8 (0.6)	0.34 (0.36)	0.8 (1.4)	1.1 (2.0)
	0d: 0a, 0e, 0i, 0g	1.06 (0.66)	0.5 (0.7)	0.7 (1.0)	0.29 (0.29)	0.7 (0.7)	1.0 (1.0)
0f,0d	0f: 0b, 0h, 0j	1.06 (0.45)	0.5 (0.5)	0.7 (0.7)	0.33 (0.64)	0.8 (1.4)	1.1 (2.1)
	0d: 0a, 0c, 0e, 0i, 0g	1.17 (0.60)	0.5 (0.6)	0.8 (0.9)	0.30 (0.27)	0.7 (0.6)	1.0 (0.9)
	Average		0.5 (0.6)	0.8 (0.8)		0.8 (1)	1.1 (1.5)

[a]The ρ-based EMR estimate is based on the haplotype estimates divided by 5 loci; the associated haplotype SE was converted to variance (squared), divided by 5 loci to produce the average per-locus variance estimate, which was then converted to the average per-locus SE (square root of variance) estimate.

Bayesian Analyses

Initial runs in Batwing (Wilson et al. 2003) indicated that mutation rate estimates (i.e., posteriors) were highly sensitive to priors (supplementary material, Supplementary Material online). However, the joint posterior distribution was relatively independent of prior distributions for other parameters. Therefore, using the dingo data set, we conducted runs under a range of mutation rate priors and in models with and without an exponential growth phase to determine which mutation rate posteriors (largely determined by the priors) produced TMRCA estimates within the credible range. Only the runs with mutation rate estimates at least as high as the moment-based estimates above produced credible estimates of TMRCA, and this was true for both demographic models (which produced similar estimates of TMRCA for all prior sets; fig. 3*A*). In the models with exponential growth, the estimates of kappa (the natural logarithm of the current-to-ancestral population size) approached zero in all runs, indicating that the more complex demographic model was not justified over the simpler constant-population size model. Therefore, we considered only the constant population size model subsequently.

The high dependence of the posterior unscaled mutation rate estimates on mutation rate priors was coupled with compensatory estimates of ancestral population size (N; not shown), such that estimates of scaled mutation rates (i.e., $2N\mu = \theta$) were relatively stable and independent of mutation rate priors. Because the priors influenced only the relative contributions of N and μ to the scaled mutation rate, θ, but not the overall value of θ itself, we used estimates of θ in conjunction with fixed assumptions about TMRCA to produce estimates of mutation rates that were independent of mutation priors and N. Calibration of these estimates to fixed TMRCAs resulted in a consistent range of estimates (fig. 3*B*). For example, runs with mutation priors corresponding to expectations ranging from 0.0001 to 0.0050 mutations per generation produced average (across loci) posterior estimates ranging 0.00052–0.00054 per year assuming TMRCA = 3,500 years and ranging 0.00037–0.00038 per year assuming TMRCA = 5,000 years (table 4).

As with the dingo data, runs with the Bali data did not produce substantial estimates of population growth (data not presented) or qualitatively affect other parameter estimates. Runs using only the Bali dog Y chromosomes that comprised

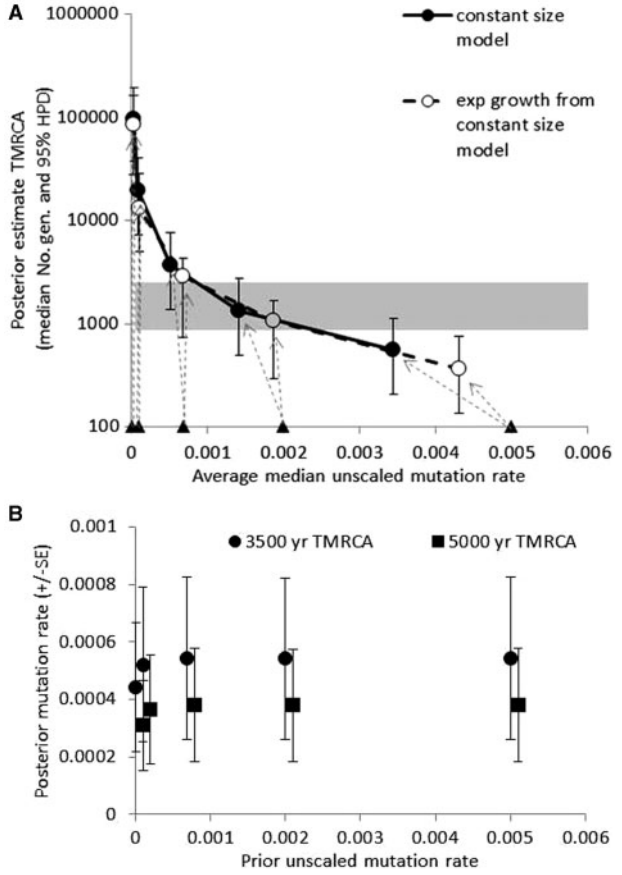

FIG. 3. Bayesian analysis of genealogical relationships among 85 dingo Y chromosomes from haplogroup H60 to produce (*A*) posterior estimates (medians and 95% HPD) of time to most recent common ancestor (TMRCA) in relation to posterior unscaled estimates of STR mutation rate (average across loci of median per-generation mutation rate) according to two demographic models and (*B*) estimates of average per-locus yearly mutation rate (±SE, $n = 5$ loci) associated with the five constant-population size runs from (*A*) corresponding to the scaled mutation rate estimates ($2N\mu$) calibrated to fixed TMRCA of 3,500 and 5,000 years versus prior mean mutation rates (per generation). Gray box in (*A*) indicates the credible range of TMRCA corresponding to 3,500 years of 4-year generations (875 generations) to 5,000 years of 2-year generations (2,500 generations); dashed gray arrows indicate posterior estimates corresponding to five sets of mutation priors (means indicated by filled triangles on axis); and the TMRCA is shown on a logarithmic scale. Symbols in (*B*) corresponding to the 5,000 year TMRCA estimates are staggered slightly along the abscissa for clarity.

haplogroup H3 (but excluding 0I, which was distantly related; supplementary fig. S1, Supplementary Material online) produced mutation rate estimates nearly identical to those for dingoes. Specifically, runs with mutation priors corresponding to expectations ranging from 0.0001 to 0.0050 mutations per generation, calibrated to TMRCA = 3,000 years, produced posterior estimates ranging 0.00055–0.00058 per year and, calibrated to TMRCA = 4,500 years, ranging 0.00037–0.00038 per year (table 4). These estimates of mutation rate in dingo and Bali dog populations were higher than the corresponding moment-based estimates above.

The locus-specific median estimates allowed calculation of SEs incorporating both sampling error and variance among loci and ranged between 0.8×10^{-4} and 2.8×10^{-4} for most prior sets (calculated from data in table 4). However, allowing loci to vary independently resulted in a complex joint posterior distribution, preventing estimation of highest posterior density (HPD) intervals for a general STR EMR. Therefore, using the dingo H60 data set, we also conducted scaled runs constraining loci to share the same mutation rate. Multiple runs using different priors produced stable posteriors. Therefore, we present results from a single set of priors (but replicated with multiple random number seeds; expected $\theta = 2.09$, gamma (1, 0.478)). The posterior estimate of median mutation rate was lower in these runs (i.e., when loci were not allowed to vary), specifically, 4.2×10^{-4} (95% HPD = 2.3×10^{-4}–6.6×10^{-4}) per year assuming a 3,500-year TMRCA and 2.9×10^{-4} (1.6×10^{-4}–4.6×10^{-4}) per year assuming a 5,000-year TMRCA.

The BATWING analysis using 195 Southeast Asian village dogs and dingoes produced estimates of split times calibrated to dingo-split times of 3,500 and 5,000 years (fig. 4). The corresponding average (across loci) mutation rate estimates were 2.4×10^{-4} (95% HPD = 1.7×10^{-4}–3.2×10^{-4}) and 1.7×10^{-4} (1.2×10^{-4}–2.3×10^{-4}) per year, which were in close agreement with those from the moment-based estimators but lower than those based on the BATWING runs calibrated to the TMRCA in dingo and Bali dog populations. Median posterior estimated split times for Bali dogs were 1,122 (95% HPD = 435–2,028) BP and 1,602 (95% HPD = 621–2,897) BP for 3,500 year and 5,000 year dingo split time calibrations, respectively, which was considerably more recent than the TMRCA estimates for this population based on the same calibrations (see earlier). Although we did not explicitly model gene flow, these results are consistent with a period of connectivity of Bali to other Southeast Asian dog populations postdating its establishment.

Although it is tempting to infer a geographic pattern of migration from the branching order reflected in figure 4, the low sample size in some of these populations and consequent overlapping HPDs warrant cautious interpretation. Nevertheless, the close phylogenetic clustering of haplotypes from Thailand, Brunei, Bali, and the Philippines suggests these populations originated from the same source, consistent with a single migration event, whereas the dingoes, NGSDs, and dogs from Taiwan appear sufficiently distinct from these to reflect a distinct migration event (fig. 2*B* and *C*). The clustering of the three Island Southeast Asian populations with Thailand also was more consistent with origination from Mainland Southeast Asia than Taiwan (in agreement with mtDNA findings of Oskarsson et al. 2011). However, the phylogenetic proximity of dingo and NGSD H60 haplotypes with Taiwanese dog H5 haplotypes suggests the possibility that the dogs of Oceana arose directly from Taiwan.

Comparison to Human STR Mutation Rates

We transformed yearly mutation rate estimates to per-generation mutation rate estimates (assuming 2- and 4-year generation times) to enable direct comparison to studies of human Y chromosome STR mutation rates. All moment-based estimates, including those assuming 2 or

Table 4. Estimates of Stepwise Mutation Rate (Per-Year) in Five Y Chromosome STR Loci Using Bayesian Analyses of Dingo ($n = 85$) and Bali Dog ($n = 37$) Genealogies Calibrated to the Insular Populations at Minimum and Maximum Times Since Descent from a Common Founding Male Ancestor (TMRCA).

Population	Prior $\bar{\mu}$ ($\times 10^{-4}$)	Assumed TMRCA (BP)	Posterior Median Mutation Rates ($\times 10^{-4}$)					
			Average	79.2	79.3	990.35	ms34CA	ms41b
Dingo	50	3,500	5.4	1.9	13.1	0.3	0.3	11.5
Dingo	20	3,500	5.4	1.9	13.0	0.3	0.3	11.5
Dingo	6.9	3,500	5.4	2.0	12.9	0.3	0.3	11.6
Dingo	1	3,500	5.2	1.9	12.5	0.3	0.3	11.0
Dingo	0.1	3,500	4.4	1.8	10.6	0.3	0.3	9.1
Dingo	50	5,000	3.8	1.3	9.2	0.2	0.2	8.1
Dingo	20	5,000	3.8	1.4	9.1	0.2	0.2	8.0
Dingo	6.9	5,000	3.8	1.4	9.1	0.2	0.2	8.1
Dingo	1	5,000	3.7	1.3	8.7	0.2	0.2	7.7
Dingo	0.1	5,000	3.1	1.3	7.4	0.2	0.2	6.4
Bali	50	3,000	5.8	4.5	10.4	4.5	2.2	7.1
Bali	20	3,000	5.7	4.6	10.2	4.5	2.2	7.2
Bali	6.9	3,000	5.7	4.6	10.2	4.5	2.2	7.2
Bali	1	3,000	5.5	4.5	9.6	4.3	2.2	6.9
Bali	0.1	3,000	4.6	3.9	7.5	3.7	2.0	5.8
Bali	50	4,500	3.8	3.0	6.9	3.0	1.5	4.8
Bali	20	4,500	3.8	3.0	6.8	3.0	1.5	4.8
Bali	6.9	4,500	3.8	3.0	6.8	3.0	1.5	4.8
Bali	1	4,500	3.7	3.0	6.4	2.9	1.4	4.6
Bali	0.1	4,500	3.0	2.6	5.0	2.5	1.3	3.9

NOTE.—Prior mutation rates followed a gamma distribution with shape parameter, 1, and scale parameter, $1/\bar{\mu}$.

4 year generations or minimum versus maximum time since founding, fell within 1 SE of the EMR for human Y STRs inferred from multiple methods (fig. 5). However, the Bayesian estimates based on TMRCA in both dingoes and Bali dogs were somewhat higher, albeit with large variability among locus estimates.

Discussion

The apparent origins of most modern dog matrilines from Southeast Asia has been interpreted as evidence that dogs were first domesticated in this region (Savolainen et al. 2002; Pang et al. 2009). However, the lack of archeological evidence of dogs in Southeast Asia until some 5,000–7,000 years later than in central and western Eurasia suggests either that the single genealogical history reflected in mtDNA could be misleading (e.g., vonHoldt et al. 2010) or that most modern dogs trace their ancestry proximately to Southeast Asia but as a secondary center of diversification associated with Neolithic rather than Paleolithic peoples (Brown et al. 2011). Previous efforts to investigate these hypotheses using independent patrilineally inherited markers have reached discrepant conclusions, owing in part to poorly resolved genealogies and unknown mutation rates (Brown et al. 2011; Ding et al. 2011). In this study, we endeavored to more finely resolve the genealogical relationship among dog patrilines (particularly those falling into the "Southeast Asian clade"; sensu Brown et al. 2011) and calibrate their EMRs (largely determined by STRs) to the known age of the Australian dingo population, enabling us to reassess the patrilineal evidence for these hypotheses. We also used other insular dog populations of Island and Mainland Southeast Asia to assist in this task and to additionally explore implications for our understanding of the Austronesian expansions.

Calibration of Y Chromosome EMRs

We used several different analytical approaches to quantify the EMRs associated with Y chromosome STRs in dogs. Most approaches produced estimates similar to those for humans on a per-generation basis. In fact, the estimates assuming a 4-year generation time in dingoes and Bali dogs yielded averages across the ρ, ASD, and BATWING split time-based estimates of 0.00069 mutations per generation, which is identical to that estimated for humans by Zhivotovsky et al. (2004) using a similar approach and similar also to the estimates by Forster et al. (2000). It is unclear why the dingo and Bali dog TMRCA-based BATWING estimates were up to 3-fold higher than the moment-based estimates of the same quantity. The larger Bayesian analysis using the entire Southeast Asian and Australian data set was more in line with moment estimated mutation rates. Moreover, these mutation rates, as calibrated to the dingo split times, were consistent with other data, for example, implying a 4,600–6,500 BP split time between

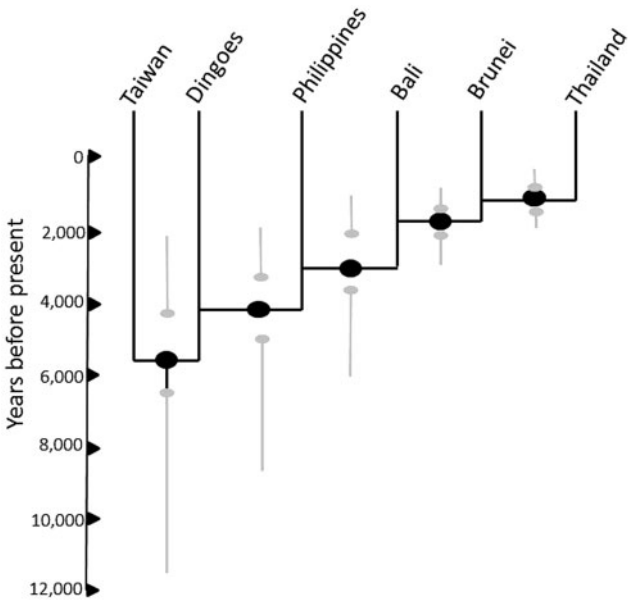

FIG. 4. Estimated splitting times among Southeast Asian village dog and dingo populations using Bayesian analysis of genealogical relationships among 195 village dog and dingo Y chromosomes (Taiwan = 4, Dingoes = 91, Philippines = 9, Bali = 63, Brunei = 11, and Thailand = 17). Branching order is tentative due to overlapping HPDs, but the pattern of haplotype clustering (fig. 2*B*) suggests that dogs from the Philippines, Bali, Brunei, and Thailand originate from a single source, potentially distinct from dingoes and dogs from Taiwan. Haplogroup H1 was excluded from this analysis. Chronology was calibrated to the dingo population assuming 3,500 year (upper gray nodes) and 5,000 year (lower gray nodes) splitting times from other Southeast Asian dogs. Black nodes are centered between these bounds with gray bars representing 95% highest posterior density (HPD) limits corresponding to 5,000 year (lower) and 3,500 year dingo splitting times (upper). These calibrations correspond, respectively, to average yearly mutation rates of 4.1×10^{-4} and 2.8×10^{-4} per locus.

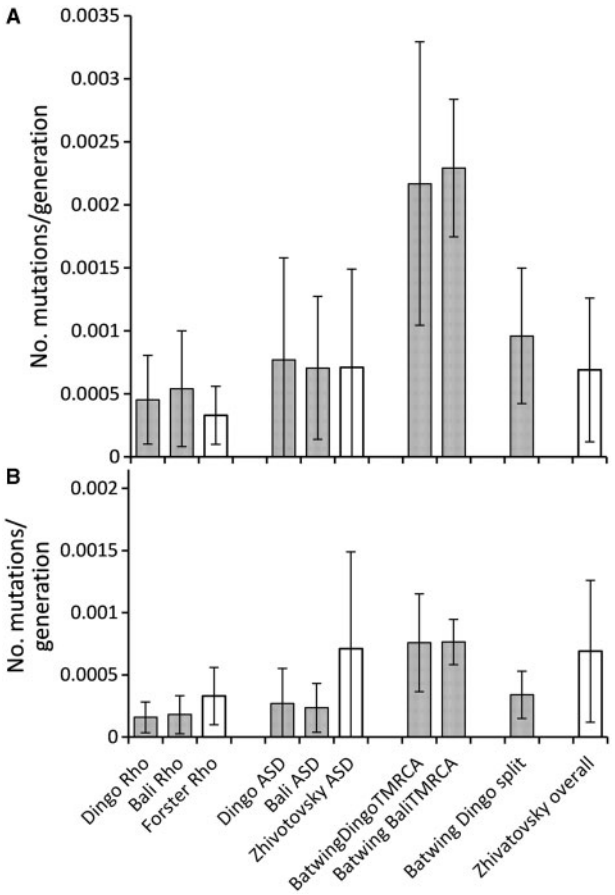

FIG. 5. Generation-scaled dog STR mutation rate estimates based on Rho, ASD, and Batwing time to most recent common ancestor (TMRCA) and split-time estimates in relation to human STR mutation rates, including a Rho estimate (Forster et al. 2000), and ASD and overall estimates (Zhivotovsky et al. 2004). Dog STR mutation rates assume (*A*) 4-year generations and 3,500- and 3,000-year isolation times for dingo and Bali dog populations, respectively, or (*B*) 2-year generations and 5,000- and 4,500-year isolation times for dingo and Bali dog populations, respectively.

Taiwanese and other Australasian dog populations, which coincided well with the time dogs were first noted via archeological remains on Taiwan and immediately preceding evidence on other islands of Southeast Asia (reviewed by Larson et al. 2012).

Regarding the use of these findings for the temporal interpretation of other studies using Y chromosome markers, the most pertinent consideration is not the absolute rate of STR (or SNP) mutation as measured from a direct accounting of mutations along a pedigree but rather the realized rate as manifested in genealogical reconstructions reflecting the evolutionary time frame of interest. One problem with assuming a single rate for all applications is that the degree of state change per unit time is time scale dependent. In both mitochondrial sequence data and nuclear STR data, pedigree and EMRs are known to differ by 10-fold or more (e.g., Forster et al. 2000; Howell et al. 2003; Zhivotovsky et al. 2004; Ho et al. 2005). Nevertheless, the relative measures of mutation rates among markers or species on one scale provide useful predictors of their relative rates on other time scales (Shi et al. 2009). In our study, we found that EMRs of Y chromosome STRs standardized to generation time were similar in dogs and humans, which was expected based on previous estimates of pedigree mutation rates. Although mutation rates of tetranucleotide STRs can be up to an order of magnitude higher in dogs (0.004–0.011 per generation) than in humans, dinucleotide repeats (which were used in this study) exhibit comparable pedigree mutation rates in dogs and humans (0.0015–0.003; Zhivotovsky et al. 2004; Irion et al. 2003; Parra et al. 2010), consistent with our estimates on an evolutionary time scale.

The utility of our estimates depends most critically on the accuracy of our assumptions about the founding time for Australian dingoes. There is little uncertainty about the more recent limit of our calibration range, 3,500 BP, as archeological dingo remains dated to 3,450 BP, 3,130 BP, and 3,270 BP have been found in sites spanning the Australian continent (reviewed by Fillios et al. 2012). The lower limit is less certain but unlikely to be more than 5,000 years. Although humans first reached Australia 50–30 Ka, the complete absence of dingo remains or faunal evidence of their presence on Tasmania indicates dogs must have arrived on Australia

after the submersion 11,000 years ago of the land bridge connecting the two land masses (Corbett 1995; Fillios et al. 2012). Moreover, extinctions of the thylacine (*Thylacinus cynocephalus*), Tasmanian devils (*Sarcophilus harrisii*), and multiple prey species from Australia beginning approximately 3,500 BP seem likely to have been caused by a recently introduced dingo (Fillios et al. 2012). There is little linguistic, genetic, or cultural indication of connectivity between Neolithic or pre-Neolithic Asian and indigenous peoples of Australia until the start of the Austronesian expansion, 6–5 Ka, after which time dingoes must have travelled to Australia with seafaring humans from the north (Corbett 1995; Hurles et al. 2003). Moreover, evidence of dogs in Taiwan or other parts of Island Southeast Asia potentially serving as source populations or routes of travel is generally lacking until 6–5 Ka if not closer to 4,500 BP (Bellwood 1997; Larson et al. 2012). Thus, the 5–3.5 Ka time window assumed in this study seems well supported by anthropological and ecological evidence.

Implications for Dating the Continental Southeast Asian Dog Expansion

Two previous studies have attempted to understand the Y chromosome phylogeography of dogs in Eurasia but neither was able to confidently time the link between east and west. By combining the SNPs of Ding et al. (2011) with the now-calibrated STRs data from Brown et al. (2011), we can re-evaluate the conclusions of these studies. Both studies concluded that the Y chromosome data were consistent with an origin of most contemporary dogs from Southeast Asia, in agreement with mtDNA studies (Savolainen et al. 2002; Pang et al. 2009; Brown et al. 2011). The fundamental difference in conclusions of Ding et al. and Brown et al. was the timing. Ding et al. concluded that a mutation (defining HG1) separating European dogs from most Southeast Asian dogs must have been ancient, predating 15,000 BP, when dogs had been hypothesized to have expanded across Eurasia (Savolainen et al. 2002; Pang et al. 2009). Brown et al. (2011), who did not assay for the HG1 mutation, concluded based on Y-STR data (and a subset of the SNPs used by Ding et al.) that the divergence between European breed dogs and Southeast Asian village dogs was too recent to reflect a single point of origin predating the earliest dog remains from the West. Rather, the common ancestry of European breed dogs and Southeast Asian indigenous dogs was interpreted to trace no further back that the Neolithic period and possibly to reflect post-Victorian use of East Asian dogs in the creation of Western Breeds.

In light of findings from this study, it seems clear that both post-Victorian and Neolithic exchanges link eastern and western Eurasian dogs. However, the cause of post-Victorian haplotype sharing between Western breed dogs and Southeast Asian village dogs apparently reflects very recent introduction of Western dogs to the East rather than extraction of Eastern dogs to create Western breeds during the Victorian Era. The best evidence for this interpretation was that haplotypes shared between Breed dogs and Southeast Asian village dogs were otherwise associated with European and American, but not Asian, breeds. These haplotypes tended also to be of the H1 haplogroup and shared among multiple Southeast Asian village dog populations, whereas other haplotypes tended to be geographically unique to populations within Southeast Asia. More importantly, this finding, afforded by combining the markers of Ding et al. and Brown et al., enabled us to age this H1 mutation according to the STR types in that haplogroup, which, in turn, provided an estimate of the divergence between European and Southeast Asian patrilines. We estimated the age of the split to fall somewhere between the more speculative estimates by Ding et al. (2011) and Brown et al. (2011).

Specifically, our aging of this European haplogroup to 5,800 (±SE = 1,750) or 8,400 (±SE = 2,500) years (depending on the dingo calibration to 3,500 or 5,000 years, respectively) suggests that the connection between pre-Victorian European and Southeast Asian dogs traces only to the Neolithic period and is not of sufficient antiquity to support the hypothesis of a single origin of dogs from Southeast Asia. Thus, although future studies are needed to combine the Y SNPs and STR markers in a geographically broader sampling of dogs than was considered here, our findings support the hypothesis for a massive Neolithic expansion of dogs from Southeast Asia rather than a Paleolithic origin of dogs from this region.

More generally, our findings that modern Eurasian continental dog populations reflect some degree of ancient admixture and that more recently founded Island populations reflect greater genetic isolation are in general agreement with the conclusions of a recent genome-wide study of Breed dogs (Larson et al. 2012). However, our findings also lead us to be far more optimistic than Larson et al. (2012) about the value of modern continental dog samples, especially nonwestern village dogs (see also Boyko et al. 2009), to inform about past migrations. For example, Brown et al. (2011) found a reciprocally monophyletic relationship between East Asian and Middle Eastern village dog patrilines, which, calibrated to mutation rate estimates from this study, were on the order of several to many thousand years, clearly indicating preservation of substantial phylogeographic signal between these continental Asian populations. This study further demonstrated the ability to use these high-resolution markers to distinguish post-Victorian gene flow from that stemming from more ancient migrations. Although the Y chromosome markers used in this study were limited in reflecting only a single (paternal) genealogy, the same principle can be applied in the future to autosomal haplotypes composed of multiple closely linked STRs and SNPs, providing replicate genealogies for statistically robust inferences.

Implications for the Austronesian Expansion

Our data also revealed aspects of the prehistory of dogs and, by extension, humans in Island Southeast Asia and Oceana. In general, our Y-chromosome aging of dog populations in

Island Southeast Asia agreed with previous estimates for timing of the Austronesian expansion. In particular, our estimated TMRCA in Bali dogs supported the arrival of their ancestors 4.5–3 Ka, coinciding with the arrival of other domesticates and elements of Neolithic culture (Bellwood 1997; Anderson 2001). Sample size was not large enough in the Philippines or Brunei to similarly age these populations nor was it clear the extent to which the haplotype clustering among these populations, Bali, and Thailand reflected colonization from a single source versus subsequent gene flow among them. The somewhat later split date than TMRCA estimated here for the Bali dogs suggests a protracted period of genetic exchange postdating their initial establishment. Nevertheless, the close clustering of the Philippines, Bali, and Brunei with Thailand but not Taiwan was more consistent with a mainland than Taiwanese origin of Island Southeast Asian dogs. The mtDNA of our sample also supported a recent common ancestry among the Philippines, Bali, Brunei, and Thailand (Brown et al. 2011). Thus, together, our Y chromosome and mtDNA findings support the general conclusion of Oskarsson et al. (2011) that dogs of Island Southeast Asia more likely originated from mainland Southeast Asia rather than Taiwan (although our findings indicated no evidence of a break between Indonesia and the Philippines as suggested in the latter study).

However, our findings also suggest that the introduction of dingoes to Australia and New Guinea could reflect a distinct event, not necessarily related in time or point of origin to the spread of dogs to Island Southeast Asia. Taiwan remains a viable candidate for the source for this particular migration. Both the haplogroup network and the Bayesian estimates of population splitting indicated a comparably ancient divergence of dingoes from the other populations. Dingoes also carry a unique haplogroup, derived from a haplogroup (H5) not sampled in Island Southeast Asia, arguing against their proximate origin from Island Southeast Asia. The H5 haplogroup was relatively common in Taiwan, however, and the STR types in that haplogroup also were highly divergent from one another, consistent with considerable antiquity relative to the shallow genealogies observed in other haplogroups and locations. One H5 haplotype from Taiwan clustered with the H60 haplogroup, differing by only two STR mutations (and the SNP), consistent with a common ancestor 4–5 Ka. On the other hand, the timing also would be consistent with origins of both the dingo (and NGSD) and Taiwanese dogs from the same expansion of Daic people of southern China (Li et al. 2008). The H5 type was reported previously in low frequency in dogs from Cambodia, China, and Siberia and in high frequency in Japan (Ding et al. 2011). Because no other domesticates or human commensals made it to ancient Australia, determining the origins of dingoes would provide a critically important piece in the puzzle regarding the interchange between Indigenous Australian and Austronesian humans preceding the peopling of Polynesia (Corbett 1995; Bellwood 1997; Oskarsson et al. 2011).

Conclusions

By using the known-age dingo population in Australia for calibration, we confirmed the EMRs of dog Y chromosome STRs to be similar on a per-generation basis (and much more rapid in absolute time) to those of humans, which have been used to date numerous late-Pleistocene and -Holocene events. Our estimates, in turn, enabled us to age the H1 mutation associated with modern European dogs and to reject the assumption that its origin predated the Neolithic. This finding argues against the hypothesis that dogs were originally domesticated in Southeast Asia and suggests, instead, that a Neolithic expansion of dogs from Southeast Asia partially replaced the earliest dogs from the West. We further hypothesize that, although dogs did not originate in Southeast Asia, they underwent a significant evolutionary transformation in this region that enabled them to demographically dominate and largely replace earlier western forms and that this transformation could have been central to the evolution dog diversity. The dogs of southern China could have been the first large population to have been reproductively isolated from wolves (*Canis lupus*), possibly accelerating their phenotypic divergence and diversification as a domesticate. Rapid spread of such a dog in combination with interbreeding with ancient western dogs could have produced a variety of forms, laying the basis for the first ancient regional breeds, which were first evident approximately 8,000 years ago (Larson et al. 2012). If so, the dingo and NGSD, which our data suggest predated the dogs of Island Southeast Asia, would reflect the last vestiges of this early Southeast Asian dog before its reacquaintance with western continental lineages.

Materials and Methods

Samples

In total, we used 338 male dogs, including 89 dingoes and 18 NGSDs genotyped for this study, along with 5 dingoes and 226 male dogs used in a previous study (Brown et al. 2011) but genotyped at an additional 18 SNP loci for this study. The 231 males selected from the previous study were those with haplotypes clustering in the "Southeast Asian clade" (sensu Brown et al. 2011), including dogs from Bali ($n = 64$), Brunei ($n = 19$), Philippines ($n = 15$), Taiwan ($n = 17$), Thailand ($n = 28$), Iran ($n = 12$), breed dogs ($n = 71$), and 5 dingoes from Fraser Island. We also included an additional five female dingoes from Fraser Island for mtDNA analysis (see later). We extracted DNA from muscle tissue of 89 new dingoes primarily from the interior of northwestern Australia, the part of the continent where genetic purity of dingoes is high (Stephens 2011). These individuals tested as "pure" based on both the "average 3Q" method (Elledge et al. 2008) and an independent method based on model-based clustering analysis of 23 microsatellite loci in >4,000 dingoes (Stephens 2011). We applied a third, confirmatory criterion in this study based on mtDNA D-loop sequences, which were diagnostic for indigenous matrilines and therefore more sensitive than the other methods for detecting historical (matrilineal) introgression (Savolainen et al. 2004; Ardalan et al. 2011).

The 18 NGSD samples were buccal swabs of individuals bred in captivity, originating from only 1–7 male founders (Koler-Matznick et al. 2003); therefore, we used them only to identify haplotypes (e.g., those in common with dingoes) but not for quantitative analyses.

Laboratory Methods

Tissue DNA was extracted according to methods of Ivanova et al. (2006), and swabs were extracted in NaOH as described previously (Brown et al. 2011). We amplified, via polymerase chain reaction (PCR), and sequenced a 582-bp fragment of the mtDNA D loop of the new dingoes using previously published primers (Savolainen et al. 2002) and sequenced in both directions with Big Dye chemistry (Applied Biosystems) using protocols described in Brown et al. (2011). We genotyped five dinucleotide-repeat STRs from the NRY, including 650-79.2b, 650-79.3b, 990-35 (Bannasch et al. 2005), MS34CA, and MS41B (Sundqvist et al. 2001) as described previously (Brown et al. 2011). We also genotyped 29 Y-chromosome SNPs using iPLEX Sequenom MassARRAY system (Sequenom Inc., San Diego, CA) using PCR and extension primers and concentrations described in supplementary table S3, Supplementary Material online. The 29 SNPs included 11 used by Brown et al. (2011) (Natanaelsson et al. 2006) in addition to 18 others that enabled differentiation of most dog haplotypes discovered in the sequencing of more than 14,000 bp of 160+ dogs and dingoes (Ding et al. 2011; Ardalan et al. 2012). The only published Y chromosome SNPs not included in our assay were Ydog-N_656, a rare polymorphism distinguishing H17 (found in a single Vietnamese dog), and Ydog-Q_773, which distinguished H2* (observed previously only in 3 European or African dogs) from H1* (Ding et al. 2011). We were not confident in the assay to consistently differentiate H2*/H1* from H1, the difference being whether a duplication producing a mutant allele or both the mutant and ancestral allele was present. Therefore, for this study, we use "H1" to refer to all three of these haplotypes (H2*, H1*, and H1; Ding et al. 2011). We successfully genotyped tissue-extracted DNA (including most of the dingo samples) in a single multiplex of all 29 loci and initially attempted to genotype swab samples in 29-locus multiplexes as well. However, many of the swab samples failed to produce full genotypes (or at least unambiguous ones) and were therefore rerun using a subset of 15 markers necessary to resolve haplotypes within the Southeast Asian clade (supplementary table S3, Supplementary Material online). All samples had been determined to be in the Southeast Asian clade based on an earlier round of genotyping, either in the study by Brown et al. (2011) or in the 29-plex attempt in this study.

Data Analysis

To illustrate the genealogical relationships among Y chromosome haplotypes and haplogroups (Y-STR haplotypes sharing the same Y-SNP haplotype), we constructed statistical parsimony networks with the median joining (MJ) algorithm implemented in Networks (v 4.6; Bandelt et al. 1999; Forster et al. 2000). Although the reduced median algorithm performs better at resolving deeper rooting lineages when only STRs are used, the MJ approach is better suited for using locus weighting information to resolve homoplasy in closely clustered haplogroups (Forster et al. 2000). Because we incorporated 29 biallelic unique event polymorphisms (SNPs) in our analysis, we were confident in the deeper phylogenetic structure (i.e., among haplogroups) of the network and therefore opted to use the MJ algorithm expected to better resolve haplotype clusters within haplogroups. As per Brown et al. (2011), we weighted the STRs inversely to their variance in repeat length and weighted SNP loci 10 times the highest STR weight (the maximum allowed by the program). Specifically, we weighted STR loci as follows: *650–79.2b* = 5, *650–79.3b* = 2, *990–35* = 9, *MS34CA* = 6, *MS41B* = 1, and SNPs were weighted 90. Otherwise, we used default settings ($r = 2$, $\epsilon = 0$).

Breed dogs were included in the first network, along with NGSDs and Iranian village dogs (the small subset from Brown et al. [2011] that had haplotypes clustering in the "Southeast Asian clade"), for reference and to explore the degree to which recent translocations of dogs might affect the various populations. A second network using only Southeast Asian village dogs and Australian dingoes was constructed and provided the basis for mutation rate estimates.

The EMR of a marker was first estimated based on the average number of mutational steps (ρ) separating modern haplotypes (derived) from the presumed founding haplotype (ancestral) divided by the time since founding (Morral et al. 1994; Forster et al. 2000). The SE of the estimate was estimated using the tree-specific algorithm of Saillard et al. (2000) in program Networks (v 4.6). Australian dingoes are known to have been established and isolated for 5,000–3,500 years (Corbett 1995; Bellwood 1997; Savolainen et al. 2004; Fillios et al. 2012). Ancestral haplotypes were identified based on internal placement and centrality in a given haplogroup within the network, frequency, and on the sharing between dingoes and NGSDs (Corbett 1995). A similar approach was used with the Bali dogs for comparison despite the less certain isolation history of the population, presuming that dogs arose with the first Austronesian colonizations 4,500–3,000 years ago (Bellwood 1997; Karafet et al. 2005).

Although this approach to estimating mutation rate is relatively robust to the existence of population structure, certain demographic scenarios result in underestimated SEs (Cox 2008). Moreover, the estimates assume particular ancestor-descendent relationships reflected in the network. Therefore, we used a complimentary approach that was less dependent upon assumptions about specific ancestor-descendent relationships among haplotypes (except for the need to assume an ancestral node). Specifically, for the main haplogroup within each of the dingo and Bali dog populations, we calculated the ASD in repeat length for all loci between a presumed ancestral haplotype and all descendent haplotypes (Zhivotovsky et al. 2004). The ASD has the expected value of the effective (i.e., as if all mutations were stepwise) mutation rate multiplied by the coalescent time of the sample (Goldstein et al. 1995). SEs were estimated as

the standard deviation (SD) divided by the square root of the number of STR loci (i.e., 5), where the SD applied to the locus-specific ASD estimates (Zhivotovsky et al. 2004).

Bayesian Analysis

To both account for uncertainty in genealogical relationships among haplotypes and allow for effects of demographic processes such as population growth, we used a Bayesian approach to sample a range of genealogical trees generated via the coalescent according to their likelihood based on the data to estimate posterior probability distributions for mutation rate (μ), TMRCA, ancestral effective male population size (N), splitting times (t) between populations, and other demographic model-dependent parameters. We conducted three sets of analyses. First dingoes, then Bali dogs, were analyzed independently, followed by a population-wide analysis of all Southeast Asian village dogs and dingoes. We conducted analyses using the Metropolis-Hastings algorithm of Wilson et al. (2003) as implemented in program BATWING, which can accommodate three demographic models: constant population size, pure exponential growth, or exponential growth from a population previously of constant size. Given that dingoes arose at least 3,500 years ago from a small number of founders, their demographic history likely included an exponential growth phase followed by slowing growth and constant size once carrying capacity was reached. Dingo remains dated across the continent by approximately 3,000 BP (Fillios et al. 2012) suggest that carrying capacity was reached long in the past and that more recent demographic history should be well approximated by a constant-size model. Nevertheless, we initially employed the most flexible model that of constant ancestral size followed by exponential increase. This model allowed the ancestral population size to range from approaching zero and growth beginning at that time (as with pure exponential growth) or approaching the current populations size with growth beginning very recently and/or a growth rate approaching zero (as with the constant population model). These runs were compared with runs of the constant-population size model to assess sensitivity of posteriors to model choice.

During initial runs to estimate μ independently of N (unscaled parameters), we found posterior estimates of μ to be highly dependent on the specified prior distribution of μ and posterior estimates of N, in turn, to be highly dependent on posterior estimates of μ (supplementary material, Supplementary Material online). Because unscaled parameters tend to be less sensitive to prior distributions (Wilson et al. 2002) and to remove the nuisance parameter (N) from computations, we ultimately used scaled estimates to test a range of mutation rates against the range of realistic TMRCA. Our approach involved the following steps: 1) conduct multiple runs with unscaled parameters, each with a different prior mutation rate distribution and the same broad prior distributions for N and demographic parameters (as applicable), 2) compare posterior estimates of mutation rate and TMRCA relative to a credible range of TMRCA based on prior information, 3) compute median posterior estimate of the scaled mutation rate, $\theta = 2N\mu$, where N and μ were the median posterior estimates of those parameters, and 4) convert posterior estimates of θ into unscaled mutation rates corresponding to the expected range of TMRCA. We performed this last step based on the relationships, $\mu = \theta/(2N)$ and TMRCA = TN. We calculated bounded estimates of the average yearly per-locus mutation rate as follows:

$$\mu_{\min} = \theta T/2(\text{Max BP})$$

and

$$\mu_{\max} = \theta T/2(\text{Min BP}),$$

where Min BP and Max BP correspond to the minimum and maximum number of years considered reasonable for the isolation time of the population (e.g., 5,000 and 3,500 years, respectively, for the dingo).

Priors and Model Parameters

We chose STR mutation rate priors following gamma distributions with shape parameter set to 1 to ensure values approaching zero were well sampled regardless of the expected value, which was controlled by varying the scale parameter. Prior ranges were chosen to include expected per-generation values estimated for human Y STRs, 0.00069 (MacPherson et al. 2004; Zhivotovsky et al. 2004; Shi et al. 2009), as well as those approximately an order of magnitude slower and faster, specifically gamma(1,200), gamma(1,500), gamma(1,1,450), gamma(1,10^5), and gamma(1,10^6). For comparison, gamma(1.5,2,175) was also used, which had the same expected mutation rate (0.00069) as gamma(1,1,450) but a higher frequency of intermediate values and was more consistent with the variability in mutation rates among human Y STRs (Zhivotovsky et al. 2004). The mutation rates were allowed to vary across loci. We used SNPs (i.e., unique event polymorphisms) to guide genealogies ("inftype = 0") but only STR mutations were used in estimates of mutation rates and time estimates. We used a prior distribution for N, gamma(1,10^{-4}), having the expected population size of 10,000, but for which the shape parameter ensured good sampling of values approaching zero. In models with population growth, we also used weakly informative prior distributions for alpha (scaled growth rate) and beta (scaled time of start of population growth), respectively, to gamma(2,400) (expected value = 0.005) and gamma(2,1) (expected value = 2) as per Xue et al. (2006). Posteriors were based on 14×10^6 Markov Chain Monte Carlo cycles (from which 70,000 samples were drawn), which was sufficient to achieve effective sample sizes (ESS) $\gg$ 200 for all likelihood and posterior estimates in single population analyses. For runs involving combined Southeast Asian village dogs and Australian dingoes (see later), the number of cycles was increased to 200×10^6 (10^6 samples drawn) to achieve ESS $\gg$ 200. Convergence was confirmed

through replicate runs initiated with different random number seeds. Likelihoods and posterior distributions were evaluated in program Tracer (v 1.5; Rambaut and Drummond 2007).

Australasian Dog Phylogeography

We took advantage of the entire Australasian dog data set to investigate splitting times of the various island populations. In particular, we tested the prediction that the splitting time estimated for dingoes in this analysis would be the same as the TMRCA of the analysis based strictly on the dingo data set, both of which we assumed would be equal to the time since establishment of that population. In addition, we assessed the extent to which the Bali dog population experienced immigration after establishment through comparison of the within-population TMRCA and the estimated splitting time. Our expectation was that if the Bali population had remained isolated since founding 4,500–3,000 years ago (Bellwood 1997; Karafet et al. 2005), the splitting time would be similar to the TMRCA estimated in the previous analysis. Parameters of analyses were the same as the previous BATWING analyses except the model incorporated multiple populations (migmodel = 1) with splitprior set to gamma (1,0.5) (expected splitting times uniformly set to 2 coalescent units) and default parameters otherwise; we used the best mutation rate prior based on calibration to the TMRCA in dingoes, and a larger prior for N, to accommodate the presumably larger size of the ancestral Australasian dog population (gamma(1,10^{-5}), expected N = 100,000).

Comparison to Human STR Rates

For comparison to mutation rates in human Y STR markers, the per-year mutation rate estimates were transformed to per-generation estimates assuming generation time ranged 2–4 years (Corbett 1995) and, conservatively, that the TMRCA for dingoes ranged 875 generations (3,500 years of 4-year generations) to 2,500 generations (5,000 years of 2-year generations). Similarly, Bali-based estimates were transformed to 750 or 2,250 generations reflecting their 4,500 to 3,000 BP founding time.

Supplementary Material

Supplementary material, tables S1–S3, and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

For samples, the authors thank the late A. Wilton, L. Allen, J. Koler-Matznik, Soi Dog Foundation (Thailand), Brunei Animal shelter, Philippine Animal Welfare Society, Veterinary Genetics Dog Research group, and Ester Honey Foundation. For technical assistance, they thank J. Malvick, T. Kun, and B. Wictum. Two anonymous reviewers provided helpful suggestions. This work was supported by the National Science Foundation (ARC-1108175), Pacific Rim Research Program (LSP#08-105) of the Center for Global, International, and Regional Studies and the Veterinary Genetics Laboratory and Center for Companion Animal Health at University of California, Davis.

References

Anderson A. 2001. Mobility models of Lapita migration. In: Clark GR, Anderson AJ, Vunidilo T, editors. The archaeology of Lapita dispersal in Oceania: papers from the Fourth Lapita Conference, June 2000, Canberra, Australia. Terra Australis 17. Canberra (Australia): Pandanus Books. p. 15–24.

Ardalan A, Kluetsch CFC, Zhang AB, Erdogan M, Uhlén M, Houshmand M, Tepeli C, Ashtiani SRM, Savolainen P. 2011. Comprehensive study of mtDNA among Southwest Asian dogs contradicts independent domestication of wolf, but implies dog–wolf hybridization. *Ecol Evol.* 1:373–385.

Ardalan A, Oskarsson M, Natanaelsson C, Wilton AN, Ahmadian A, Savolainen P. 2012. Narrow genetic basis for the Australian dingo confirmed through analysis of paternal ancestry. *Genetica* 140: 65–73.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.

Bannasch D, Bannasch M, Ryun J, Famula T, Pedersen N. 2005. Y chromosome haplotype analysis in purebred dogs. *Mamm Genome.* 16: 273–280.

Bellwood P. 1997. Prehistory of the Indo-Malaysian archipelago. Honolulu (Hawaii): University of Hawaii Press.

Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhano M, et al. (15 co-authors). 2009. Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc Natl Acad Sci U S A.* 106:13903–13908.

Brown SK, Darwent CM, Sacks BN. 2013. Ancient DNA evidence for genetic continuity in Arctic Dogs. *J Archaeol Sci.* 40:1279–1288.

Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, Wu JT, Okon M, Sacks BN. 2011. Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. *PLoS One* 6:e28496.

Bulmer S. 2001. Lapita dogs and singing dogs and the history of the dog in New Guinea. In: Clark GR, Anderson AJ, Vunidilo T, editors. The archaeology of Lapita dispersal in Oceania: papers from the Fourth Lapita Conference, June 2000, Canberra, Australia. Terra Australis 17. Canberra (Australia): Pandanus Books. p. 183–201.

Clutton-Brock J. 1995. Origins of the dog: domestication and early evolution. In: Serpell J, editor. The domestic dog: its evolution, behaviour and interaction with people. Cambridge: Cambridge University Press. p. 8–20.

Corbett LK. 1995. The dingo in Australia and Asia. Ithaca (NY): Cornell University Press.

Cox MP. 2008. Accuracy of molecular dating with the Rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol.* 80:335–357.

Crockford SJ, Kuzmin YV. 2012. Comments on Germonpré et al., Journal of Archaeological Science 36, 2009 "Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes," and Germonpré, Lázkicková-Galetová, and Sablin, Journal of Archaeological Science 39, 2012 "Palaeolithic dog skulls at the Gravettian Predmostí site, the Czech Republic.". *J Archaeol Sci.* 39:2797–2801.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.

Dikov NN. 1996. The Ushki sites, Kamchatka Peninsula. In: West FH, editor. American beginnings: the prehistory and palaeoecology of Beringia. Chicago (IL): University of Chicago Press. p. 244–250.

Ding ZL, Oskarsson M, Ardalan A, Angleby H, Dahlgren LG, Tepeli C, Kirkness E, Savolainen P, Zhang YP. 2011. Origins of domestic dog in

Southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* 108:507–514.
Dobney K, Larson G. 2006. Genetics and animal domestication: new windows on an elusive process. *J Zool.* 269:261–271.
Elledge AE, Leung LKP, Allen LR, Firestone K, Wilton AN. 2008. An evaluation of genetic analyses, skull morphology and visual appearance for assessing dingo purity: implications for dingo conservation. *Wildl Res.* 35:812–820.
Fillios M, Crowther MS, Letnic M. 2012. The impact of the dingo on the thylacine in Holocene Australia. *World Archaeol.* 44:118–134.
Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat based phylogeny for the human Y chromosome. *Am J Hum Genet.* 67:182–196.
Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471.
Henn BM, Gignoux C, Lin AA, Oefner PJ, Shen P, Scozzari R, Cruciani F, Tishkoff SA, Mountain JL, Underhill PA. 2008. Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc Natl Acad Sci U S A.* 105:10693–10698.
Heyer E, Puymirat J, Dietjes P, Bakker E, de Knijff P. 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet.* 6:799–803.
Ho SY, Phillips MJ, Drummond AJ, Cooper A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Mol Biol Evol.* 22:1355–1363.
Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnball DM, Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet.* 72:659–670.
Hurles ME, Matisoo-Smith E, Gray RD, Penny D. 2003. Untangling oceanic settlement: the edge of the knowable. *Trends Ecol Evol.* 18:531–540.
Irion DN, Schaffer AL, Famula TR, Eggleston ML, Hughs SS, Pedersen NC. 2003. Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. *J Hered.* 94:81–87.
Irion DN, Schaffer A, Grant S, Wilton A, Pedersen N. 2005. Genetic variation analysis of the Bali street dog using microsatellites. *BMC Genet.* 6:6.
Ivanova NV, Dewaard JR, Hebert PDN. 2006. An inexpensive automation-friendly protocol for recovering high-quality DNA. *Mol Ecol Notes.* 6:998–1002.
Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet.* 4:598–612.
Karafet TM, Lansing JS, Redd AJ, et al. (11 co-authors). 2005. Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum Biol.* 77:93–114.
Kayser M, Roewer L, Hedman M, et al. (13 co-authors). 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet.* 66: 1580–1588.
Koler-Matznick J, Brisbin IL, Feinstein M, Bulmer S. 2003. An updated description of the New Guinea singing dog (*Canis hallstromi*, Troughton 1957). *J Zool.* 261:109–118.
Larson G, Karlsson EK, Perri A, et al. (20 co-authors). 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A.* 109:8878–8883.
Larson G, Liuc R, Zhaoc X, et al. (14 co-authors). 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Nat Acad Sci U S A.* 107:7686–7691.
Leonard J, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C. 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* 298:1613–1616.
Li H, Wen B, Chen S-J, et al. (19 co-authors). 2008. Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol.* 8:146.
Macpherson JM, Ramachandran S, Diamond L, Feldman MW. 2004. Demographic estimates from Y chromosome microsatellite polymorphisms: analysis of a worldwide sample. *Hum Genomics.* 1: 345–354.
Matisoo-Smith E. 2007. Animal translocations, genetic variation and the human settlement of the Pacific. In: Friedlaender JS, editor. Genes, language and culture history in the southwest Pacific. Oxford: Oxford University Press. p. 157–170.
Morey DF. 2010. Dogs: domestication and the development of a social bond. New York: Cambridge University Press.
Morral N, Bertranpetit J, Estivill X, et al. (31 co-authors). 1994. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet.* 7:169–175.
Natanaelsson C, Oskarsson M, Angleby H, Lundeberg J, Kirkness E, Savolainen P. 2006. Dog Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC Genet.* 7:45.
Olsen SJ. 1985. Origins of the domestic dog: the fossil record. Tucson (AZ): University of Arizona Press.
Oskarsson MCR, Klütsch CFC, Boonyaprakob U, Wilton A, Tanabe Y, Savolainen P. 2011. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc R Soc B: Biol Sci.* 279:967–974.
Pang J-F, Klütsch C, Zou X-J, et al. (14 co-authors). 2009. mtDNA data indicates a single origin for dogs south of Yangtze river, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol.* 26: 2849–2864.
Parra D, Garcia D, Mendez S, Cañon J, Dunner S. 2010. High mutation rates in canine tetranucleotide microsatellites: too much risk for genetic compatibility purposes? *Open Forens Sci J.* 3:9–13.
Pedersen NC, Liu H, Theilen GH, Sacks BN. 2013. The effects of dog breed development on genetic diversity and the relative influences of performance and conformation breeding. *J Anim Breed Genet.* In press.
Pionnier-Capitan M, Bemilli C, Bodu P, Célérier G, Ferrié J-G, Fosse P, Garcià M, Vigne J-D. 2011. New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe. *J Archaeol Sci.* 38: 2123–2140.
Rambaut A, Drummond AJ. 2007. Tracer v1.4: MCMC trace analyses tool. Available from: http://beast.bio.ed.ac.uk/Tracer (last accessed February 22, 2013).
Sacks BN, Bannasch DL, Chomel BB, Ernest HB. 2008. Coyotes demonstrate how habitat specialization by individuals of a generalist species can diversify populations in a heterogeneous ecoregion. *Mol Biol Evol.* 25:1384–1394.
Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. 2000. mtDNA Variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 67:718–726.
Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E, Lundeberg J. 2004. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc Natl Acad Sci U S A.* 101: 12387–12390.
Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610–1613.
Shi W, Ayub Q, Vermeulen M, Shao R-G, Zuniga S, van der Gaag K, de Knijff P, Kayser M, Xue Y, Tyler-Smith C. 2009. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol.* 27:385–393.
Stephens D. 2011. The molecular ecology of Australian wild dogs: hybridization, gene flow and genetic structure at multiple geographic scales [PhD Dissertation]. University of Western Australia.
Sundqvist A-K, Ellegren H, Olivier M, Vila C. 2001. Y chromosome haplotyping in Scandinavian wolves (*Canis lupus*) based on microsatellite markers. *Mol Ecol.* 10:1959–1966.
Tanabe Y. 2006. Phylogenetic studies of dogs with emphasis on Japanese and Asian breeds. *Proc Jpn Acad, Ser B.* 82:375–387.

Tchernov E, Valla FF. 1997. Two New Dogs, and Other Natufian Dogs, from the Southern Levant. *J Archaeol Sci.* 24:65–95.

Underhill AP. 1997. Current issues in Chinese Neolithic archaeology. *J World Prehistory.* 11:103–160.

vonHoldt BM, Pollinger JP, Lohmueller KE, et al. (36 co-authors). 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898–902.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Hum Mol Genet.* 2:1123–1128.

Wilson I, Balding D, Weale M. 2002. Batwing user guide: Bayesian analysis of trees with internal node generation. Available from: http://www.mas.ncl.ac.uk/~nijw/batwing/batguide.pdf (last accessed February 22, 2013).

Wilson I, Weale M, Balding D. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc Ser A.* 166:155–188.

Xue Y, Zerjal T, Bao W, et al. (11 co-authors). 2006. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 172:2431–2439.

Zeder MA, Emshwiller E, Smith BD, Bradley DG. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22:139–155.

Zhivotovsky LA, Underhill PA, Cinnioğlu C, et al. (17 co-authors). 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 74:50–61.