High-Resolution Phylogenetic Analysis of Southeastern Europe Traces Major Episodes of Paternal Gene Flow Among Slavic Populations

Marijana Peričić,[*][1] *Lovorka Barać Lauc,*[*][1] *Irena Martinović Klarić,*[*] *Siiri Rootsi,*[†]
Branka Janićijević,[*] *Igor Rudan,*[‡][§] *Rifet Terzić,*[||] *Ivanka Čolak,*[¶] *Ante Kvesić,*[¶] *Dan Popović,*[*]
Ana Šijački,[#] *Ibrahim Behluli,*[**] *Dobrivoje Đorđević,*[††] *Ljudmila Efremovska,*[††] *Đorđe D. Bajec,*[#]
Branislav D. Stefanović,[#] *Richard Villems,*[†] *and Pavao Rudan*[*]

[*]Institute for Anthropological Research, Amruševa 8, 10000 Zagreb, Croatia; [†]Estonian Biocentre, University of Tartu, Tartu, Estonia; [‡]School of Public Health Andrija Štampar, University of Zagreb Medical School, Zagreb, Croatia; [§]University of Edinburgh Medical School, Edinburgh, Scotland; [||]Medical Faculty, University of Tuzla, Tuzla, Bosnia and Herzegovina; [¶]Clinical Hospital Center "Bijeli Brijeg," Mostar, Bosnia and Herzegovina; [#]Emergency Unit of Clinical Center of Serbia, Belgrade, Serbia and Montenegro; [**]Medical Faculty, University of Prishtina, Prishtina, Kosovo; and [††]Medical Faculty, University of Skopje, Skopje, Macedonia

The extent and nature of southeastern Europe (SEE) paternal genetic contribution to the European genetic landscape were explored based on a high-resolution Y chromosome analysis involving 681 males from seven populations in the region. Paternal lineages present in SEE were compared with previously published data from 81 western Eurasian populations and 5,017 Y chromosome samples. The finding that five major haplogroups (E3b1, I1b* (xM26), J2, R1a, and R1b) comprise more than 70% of SEE total genetic variation is consistent with the typical European Y chromosome gene pool. However, distribution of major Y chromosomal lineages and estimated expansion signals clarify the specific role of this region in structuring of European, and particularly Slavic, paternal genetic heritage. Contemporary Slavic paternal gene pool, mostly characterized by the predominance of R1a and I1b* (xM26) and scarcity of E3b1 lineages, is a result of two major prehistoric gene flows with opposite directions: the post-Last Glacial Maximum R1a expansion from east to west, the Younger Dryas-Holocene I1b* (xM26) diffusion out of SEE in addition to subsequent R1a and I1b* (xM26) putative gene flows between eastern Europe and SEE, and a rather weak extent of E3b1 diffusion toward regions nowadays occupied by Slavic-speaking populations.

Introduction

Southeastern Europe (SEE) has traditionally been viewed as a "bridge" (Childe 1958) between the Near East and temperate Europe or as a key area in the process of transition from hunter-gathering to agropastoral, farming societies in Europe (e.g., Ammerman and Cavalli-Sforza 1984; Renfrew 1987; Zvelebil and Lillie 2000). Recent phylogeographic analyses of Y chromosome E and J haplogroups indicate that southern Europe and the Balkans indeed could have been both the receptors and sources of gene flow during and after the Neolithic (Cruciani et al. 2004; Semino et al. 2004). The STR haplotype diversity of these two haplogroups is considerably younger than that of other Y chromosome haplogroups spread in Europe. Among the latter, haplogroup I, perhaps, most clearly represents the paternal genetic component of the pre-Neolithic Europeans. In contrast to E and J, haplogroup I is virtually absent in Middle East and West Asia (Semino et al. 2000), and two of its major subclades have frequency peaks in northern Balkans and Scandinavia (Rootsi et al. 2004). Semino et al. (2000) and Barać et al. (2003) hypothesized that, besides southwest Europe, the northern Balkans could have been another possible Last Glacial Maximum (LGM) refugium and a reservoir of M170.

In this study we first examined the extent and nature of SEE paternal genetic contribution to the European genetic landscape based on a high-resolution Y chromosome typing involving 681 unrelated males from four modern states,

Croatia, Bosnia and Herzegovina, Serbia and Montenegro (including the province of Kosovo), and Macedonia (fig. 1). Second, we exploited available data on Y chromosome variation among different southern, western, and eastern Slavic-speaking populations in Europe to draw conclusions about possible origin of major paternal lineages in the Slavic gene pool. Finally, based on geography, we assessed patterns of Y chromosome diversity across SEE.

Materials and Methods

We analyzed 681 males from seven populations from SEE and 5,017 Y chromosomes from 81 western Eurasian populations available from literature. Blood samples were collected from healthy unrelated adults after obtaining informed consent. DNA was extracted using the salting-out procedure (Miller, Dykes, and Polesky 1988).

The following set of biallelic markers was analyzed using restriction fragment length polymorphism (RFLP) or in/del assays according to published protocols: M9 (Whitfield, Sulston, and Goodfellow 1995), YAP (Hammer and Horai 1995), SRY-1523 (Whitfield, Sulston, and Goodfellow 1995) (SRY-1523 is equivalent to SRY_{10831} [Whitfield, Sulston, and Goodfellow 1995]), 92R7 (Mathias, Bayés, and Tyler-Smith 1994), 12f2 (Rosser et al. 2000), M170, M173, M89 (Underhill et al. 2000), and P37 (Y Chromosome Consortium 2002). The polymorphic single nucleotide polymorphism (SNP) underlying markers M26, M35, M67, M69, M78, M81, M82, M92, M102, M123, M172, M201 (Underhill et al. 2000), M223 (Underhill et al. 2001), M241, M242, M253 (Cinnioğlu et al. 2004), and $SRY_{8299/4064}$ (Whitfield, Sulston, and Goodfellow 1995) were sequenced after polymerase chain reaction (PCR)

FIG. 1.—Map of the studied region and sample locations (1 = Zabok, 2 = Zagreb, 3 = Donji Miholjac, 4 = Delnice, 5 = Pazin, 6 = Dubrovnik, 7 = Zenica, 8 = Mostar, 9 = Široki Brijeg, 10 = Belgrade, 11 = Pristhina, 12 = Skopje).

amplification. PCR-amplified products were purified using shrimp alkaline phosphatase and exonuclease treatment following Kaessmann et al. (1999) and sequenced using the BigDye Terminator Version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.) on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) by using the DNA Sequencing Analysis Software Version 3.7 (Applied Biosystems). M9 was typed on all samples, and other markers were typed hierarchically according to their known phylogeny. A tentative assignment of all R1 chromosomes derived at M173 but without the G to A back mutation at SRY_{10831} into haplogroup R1b was based on the observations of Cruciani et al. (2002). Phylogenetic relationships of analyzed biallelic markers are presented in figure 2. Mutation labeling follows the Y Chromosome Consortium (2002).

In addition, we surveyed eight short tandem repeats (STRs) DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393 (Kayser et al. 1997) on all 681 SEE chromosomes and one additional GATA STR A7.1 (DYS460) (White et al. 1999) in E3b1-M78 chromosomes. PCR products were detected on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems), and fragment sizes were analyzed by the GeneScan Analysis Software Version 3.7 (Applied Biosystems).

Expansion ranges were expressed as the age of STR variation estimated as the average squared difference in

the number of repeats of seven STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) between all sampled chromosomes and the founder haplotype divided by w (effective mutation rate of 0.00069 per locus per 25 years) (Zhivotovsky et al. 2004). Phylogenetic networks were obtained by using the same seven STRs as those used for expansion range estimates. The phylogenetic relationships between microsatellite haplotypes were determined by using the program NETWORK 4.0b (Fluxus Engineering). Networks were calculated by the median-joining method (Bandelt, Forster, and Röhl 1999), and STR loci were weighted according to Helgason et al. (2000). Haplogroup-frequency and haplogroup-variance surfaces were reconstructed following the Kringing procedure by use of the Surfer System (Golden Software), the frequency data reported in table 1 and variance data from this study and literature, as specified in figures 3–7. Credible regions (95% CRs) for haplogroup frequencies were calculated from posterior distribution of the proportion of the group of lineages in the population, as in Richards et al. (2000). For the purpose of correlating Y chromosomal frequencies with geography, we used Spearman's bivariate correlation procedure (SPSS for Windows, 7.5.1.). Sampled individuals were pooled into 12 regional towns (fig. 1) with following latitude (N) and longitude (E) values: (1) 46°02′, 15°90′; (2) 45°82′, 15°98′; (3) 45°77′, 18°17′; (4)
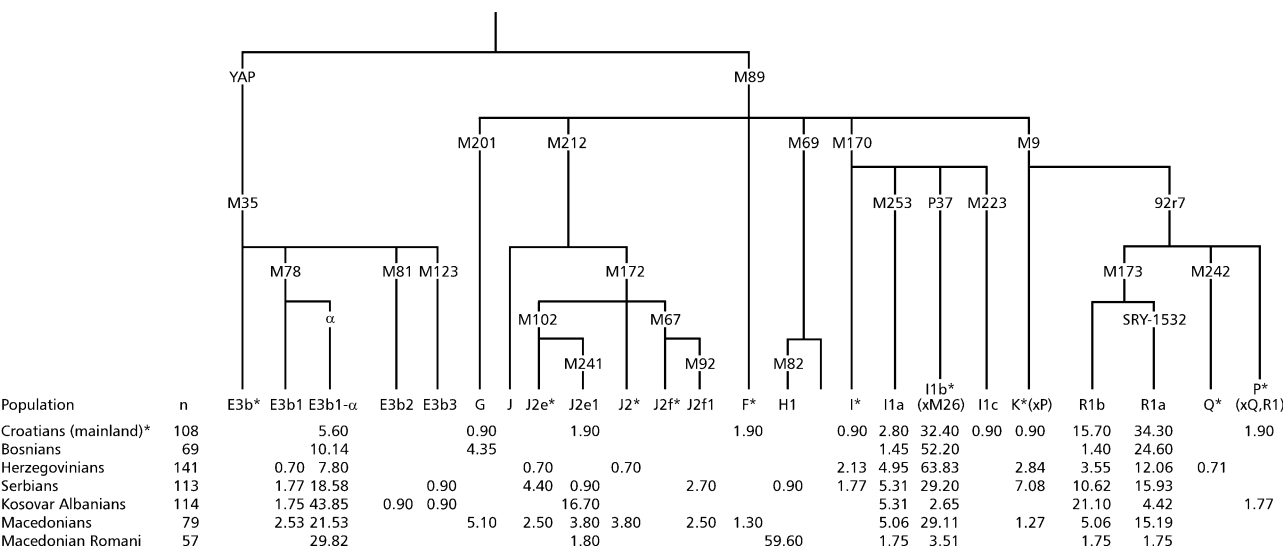
FIG. 2.—Y chromosomal SNP tree and haplogroup frequencies (percent) in seven SEE populations. *Croatian mainland from Barać et al. (2003) was additionally genotyped for deeper resolution of I in Rootsi et al. (2004) and for E and J in the present study. E3b1α chromosomes were defined by A7.1 nine-repeat allele.

Population	n	E3b*	E3b1	E3b1-α	E3b2	E3b3	G	J	J2e*	J2e1	J2*	J2f*	J2f1	F*	H1	I*	I1a	I1b*(xM26)	I1c	K*(xP)	R1b	R1a	Q*	P*(xQ,R1)
Croatians (mainland)*	108		5.60				0.90			1.90				1.90		0.90	2.80	32.40	0.90	0.90	15.70	34.30		1.90
Bosnians	69		10.14					4.35									1.45	52.20			1.40	24.60		
Herzegovinians	141	0.70	7.80							0.70	0.70					2.13	4.95	63.83		2.84	3.55	12.06	0.71	
Serbians	113	1.77	18.58			0.90				4.40	0.90		2.70		0.90	1.77	5.31	29.20		7.08	10.62	15.93		
Kosovar Albanians	114	1.75	43.85		0.90	0.90				16.70							5.31	2.65			21.10	4.42		1.77
Macedonians	79	2.53	21.53				5.10		2.50	3.80	3.80		2.50			1.30	5.06	29.11		1.27	5.06	15.19		
Macedonian Romani	57		29.82							1.80					59.60		1.75	3.51			1.75	1.75		

45°40′, 14°80′; (5) 45°23′, 13°93′; (6) 42°65′, 18°09′; (7) 44°22′, 17°90′; (8) 43°35′, 17°80′; (9) 43°39′, 17°55′; (10) 44°82′, 20°46′; (11) 42°67′, 21°17′; and (12) 41°98′, 21°43′.

Results and Discussion

One-third of the studied SEE Y chromosomes has the derived P37 C allele and is classified to haplogroup I1b* (xM26) (fig. 2). A detailed survey demonstrates that I1b* (xM26) lineages reach maximum frequency in SEE (fig. 3C) and that I1b* (xM26) STR variance peaks over a large geographic region encompassing both southeastern and central Europe (fig. 3D). I1b* (xM26) frequency peaks in Herzegovinians (64%) and Bosnians (52%) while preserving substantial (30%) frequencies in all SEE populations with the exception of two reproductively isolated and non-slavic speaking populations, Kosovar Albanians and Macedonian Romani (fig. 3A). The incidence of I1b* (xM26) decreases from SEE toward western (from 20% in Slovenians abruptly to 1% in northern Italians) and southern (17%–18% in Albanians and northern Greeks, 8% in southern Greeks, 2% in Turks) and retains frequencies of 7%–22% in central and eastern Europe (table 1). The highest STR variance of I1b* (xM26) lineages (0.34 to 0.23) is in Bosnians, Czechs and Slovaks, Hungarians, Herzegovinians, and Serbians (fig. 3B and D). In both cases, when all studied SEE populations are considered together and upon exclusion of Kosovar Albanians and Macedonian Romani, I1b* (xM26) frequency and variance do not show significant correlations with geography (table 2). Moreover, I1b* (xM26) phylogenetic network (fig. 8A) shows high haplotype diversity and sharing of founder haplotype among investigated populations. In fact, homogenous distribution of elevated frequency accompanied with high diversity of I1b* (xM26) lineages among different SEE populations may be viewed as a genetic signature of their common paternal history over a long period of time. Rootsi et al. (2004) estimated that I1b* (xM26) diverged from I* at 10.7 ± 4.8 kilo years ago (KYA), possibly relating to the post–Younger Dryas (YD) climate amelioration in Europe, and that I1b* (xM26) expansion occurred around the early Holocene at 7.6 ± 2.7 KYA. Considering only our SEE sample, the coalescent estimate of I1b* (xM26) is substantially older (11.1 ± 4.8 KYA). This finding suggests that the I1b* (xM26) lineages might have expanded from SEE to central, eastern, and southern Europe, presumably not earlier than the YD to Holocene transition and not later than the early Neolithic.

Haplogroup E3b1-M78 is the second most prevailing one (23%) in the studied sample with E3b1-M78 chromosomes accounting for almost all E representatives (98%) except a single E3b2-M81 and two E3b3-M123 chromosomes (fig. 2). E3b1-M78 is the most common haplogroup E lineage in Europe (Cruciani et al. 2004; Semino et al. 2004). The spatial pattern shown in figure 4(C) depicts a nonuniform E3b1 geographic distribution with a frequency peak centered in south Europe and SEE (13%–16% in southern Italians and 17%–27% in the Balkans). Declining frequencies are evident toward western (10% in northern and central Italians), central, and eastern Europe (from 4% to 10% in Polish, Russians, mainland Croatians, Ukrainians, Hungarians, Herzegovinians, and Bosnians). Noteworthy is a low E3b1 frequency (5%) in Turkey. Apart from its presence in Europe and the Middle East, E3b1 is also found in eastern and northern Africa. Cruciani et al. (2004) estimated that E3b-M78 might have originated in eastern Africa about 23.2 KYA (95% confidence interval [CI] 21.1–25.4). Although present level of phylogenetic resolution does not allow further subdivision of this haplogroup by binary markers, based on strong geographic structuring of diverse microsatellite motifs, E3b-M78 is suggested to be a collection of subclades with different evolutionary histories (Cruciani et al. 2004; Semino et al. 2004)

Table 1
Summarized Percent Frequencies of R1b, R1a, I1b* (xM26), E3b1 and J2e

Population	Language Family	N	R1b-M173[a]	95% CR	N	R1a-M17	95% CR	N	I1b* (xM26)-P37	95% CR	N	E3b1-M78	95% CR	N	J2e-M102	95% CR
Slovenian	IE (Balto-Slavic)	70	21.0[b]	(13.5–32.4)	70	37.0[b]	(26.8–48.9)	55	20.0[c]	(11.6–32.4)	—	—	—	—	—	—
Croatian (mainland)	IE (Balto-Slavic)	108	15.7[d]	(10.1–23.8)	108	34.3[d]	(26.0–43.6)	108	32.4[c–e]	(24.3–41.7)	108	5.6[d]	(2.6–11.6)	108	1.0[d]	(0.2–5.0)
Bosnian	IE (Balto-Slavic)	69	1.4[d]	(0.3–7.7)	69	24.6[d]	(16.0–36.0)	69	52.2[d]	(40.6–63.6)	69	10.1[d]	(5.1–19.5)	69	0[d]	(0.0–4.2)
Herzegovinian	IE (Balto-Slavic)	141	3.6[d]	(1.6–8.0)	141	12.1[d]	(7.7–18.5)	141	63.8[d]	(55.6–71.3)	141	8.5[d]	(5.0–14.3)	141	0.7[d]	(0.2–3.9)
Serbian	IE (Balto-Slavic)	113	10.6[d]	(6.2–17.7)	113	15.9[d]	(10.3–23.8)	113	29.2[d]	(21.6–38.2)	113	20.4[d]	(14.0–28.7)	113	5.3[d]	(2.5–11.1)
Macedonian	IE (Balto-Slavic)	79	5.1[d]	(0.21–12.3)	79	15.2[d]	(8.9–24.7)	79	29.1[d]	(20.3–40.0)	79	24.1[d]	(16.0–34.6)	79	6.3[d]	(2.8–14.0)
Macedonian (Greek)	IE (Greek)	20	10.0[f]	(3.0–30.4)	20	35.0[f]	(18.1–57.0)	50	18.0[c]	(9.8–30.9)	59	18.6[g]	(10.8–30.4)	—	—	—
Macedonian Romani	IE (Indo-Iranian)	57	1.8[d]	(0.4–9.2)	57	1.8[d]	(0.4–9.2)	57	3.5[d]	(1.1–11.9)	57	29.8[d]	(19.5–42.7)	57	1.8[d]	(0.4–9.2)
Albanian (Kosovar)	IE (Albanian)	114	21.1[d]	(14.6–29.4)	114	4.4[d]	(1.9–9.9)	114	2.7[d]	(1.0–7.4)	114	45.6[d]	(36.8–54.8)	114	16.7[d]	(11.0–24.6)
Albanian	IE (Albanian)	51	17.6[f]	(9.6–30.3)	51	9.8[f]	(4.4–21.0)	106	17.0[c]	(11.0–25.3)	63	27.0[g,h]	(17.6–39.1)	56	14.3[g]	(7.5–25.8)
Greek	IE (Greek)	118	22.8[f,i]	(16.2–31.3)	118	8.3[f,i]	(4.7–14.9)	261	8.4[c]	(5.6–12.4)	84	21.4[g]	(14.0–31.4)	92	6.5[g]	(3.1–13.5)
Cypriot	IE (Greek)	45	9.0[b]	(3.6–20.8)	45	2.0[b]	(0.5–11.5)	—	—	—	116	20.7[h]	(14.3–29.0)	—	—	—
Bulgarian	IE (Balto-Slavic)	24	17.0[b]	(6.8–36.1)	34	14.7[j]	(6.6–30.3)	—	—	—	—	—	—	—	—	—
Romanian	IE (Italic)	45	18.0[b]	(9.4–31.4)	45	20.0[b]	(10.9–33.9)	361	17.7[c]	(14.1–22.0)	14	21.4[h]	(7.8–48.1)	—	—	—
Hungarian	Uralic (Finno-Ugric)	45	13.3[f]	(6.3–26.3)	113	20.4[k]	(14.0–28.7)	162	11.1[c]	(7.2–16.9)	53	7.5[g]	(3.1–17.9)	49	0[g]	(0.0–5.8)
Czech and Slovak	IE (Balto-Slavic)	45	35.6[f]	(23.2–50.2)	45	26.7[f]	(16.0–41.1)	198	7.1[c]	(4.3–11.5)	—	—	—	—	—	—
Polish	IE (Balto-Slavic)	55	16.4[f]	(8.9–28.3)	55	56.4[f]	(43.2–68.7)	191	9.9[c]	(6.5–15.0)	137	3.6[g,h]	(1.6–8.3)	97	1.0[g]	(0.2–5.6)
Byelorussian	IE (Balto-Slavic)	41	10.0[b]	(4.0–22.6)	41	39.0[b]	(25.6–54.4)	147	15.0[c]	(10.1–21.6)	—	—	—	—	—	—
Russian	IE (Balto-Slavic)	122	7.0[b]	(14.2–44.6)	122	47.0[b]	(38.1–55.6)	—	—	—	—	—	—	—	—	—
Russian (Adygea)	IE (Balto-Slavic)	—	—	—	—	—	—	78	16.7[c]	(10.0–26.5)	—	—	—	—	—	—
Russian (Bashkortostan)	IE (Balto-Slavic)	—	—	—	—	—	—	50	2.0[c]	(0.5–10.4)	—	—	—	—	—	—
Russian (Belgorod region)	IE (Balto-Slavic)	—	—	—	—	—	—	144	12.5[c]	(8.1–18.9)	—	—	—	—	—	—
Russian (Cossacks)	IE (Balto-Slavic)	—	—	—	—	—	—	97	15.5[c]	(9.6–24.0)	—	—	—	—	—	—
Russian (Kostroma region)	IE (Balto-Slavic)	—	—	—	—	—	—	53	9.4[c]	(4.2–20.3)	—	—	—	—	—	—
Russian (North, Pinega)	IE (Balto-Slavic)	—	—	—	—	—	—	127	3.9[c]	(1.7–8.9)	—	—	—	—	—	—
Russian (Smolensk region)	IE (Balto-Slavic)	—	—	—	—	—	—	120	9.1[c]	(5.2–15.7)	—	—	—	—	—	—
Ukrainian	IE (Balto-Slavic)	50	2.0[f]	(0.5–10.4)	50	54.0[f]	(40.3–67.1)	585	16.1[c]	(13.3–19.3)	93	7.5[g]	(3.7–14.7)	82	2.4[g]	(0.8–8.4)
Icelander	IE (Germanic)	181	41.4[i]	(34.5–48.7)	181	23.8[i]	(18.1–30.5)	—	—	—	—	—	—	—	—	—
Swedish	IE (Germanic)	110	20.0[i]	(13.6–28.5)	110	17.3[i]	(11.4–25.4)	225	0[c]	(0.0–1.3)	—	—	—	—	—	—
Swedish (Northern)	IE (Germanic)	48	23.0[b]	(13.3–36.6)	48	19.0[b]	(10.2–32.0)	—	—	—	—	—	—	—	—	—
Saami	Uralic (Finno-Ugric)	127	3.9[k]	(1.7–8.9)	127	11.0[k]	(6.7–17.7)	35	0[c]	(0.0–8.0)	—	—	—	—	—	—
Finnish	Uralic (Finno-Ugric)	57	2.0[b]	(0.4–9.2)	57	10.0[b]	(2.9–16.7)	—	—	—	—	—	—	—	—	—
Norwegian	IE (Germanic)	112	25.9[i]	(18.7–34.7)	112	17.9[i]	(11.9–26.0)	72	0[c]	(0.0–4.0)	—	—	—	—	—	—
Danish	IE (Germanic)	12	41.7[i]	(19.2–68.4)	12	16.7[i]	(5.0–45.4)	—	—	—	35	2.9[h]	(0.7–14.5)	—	—	—
Lithuanian	IE (Balto-Slavic)	38	5.0[b]	(1.6–17.3)	38	34.0[b]	(21.2–50.2)	—	—	—	—	—	—	—	—	—
Latvian	IE (Balto-Slavic)	34	15.0[b]	(6.6–30.3)	34	41.0[b]	(26.3–57.9)	86	1.2[c]	(0.3–6.2)	—	—	—	—	—	—
Estonian	Uralic (Finno-Ugric)	207	9.0[b]	(6.0–13.9)	74	36.5[j]	(26.4–47.9)	210	2.9[c]	(1.3–6.1)	74	4.1[h]	(1.5–11.2)	—	—	—
Moldavian	IE (Italic)	—	—	—	—	—	—	60	21.7[c]	(13.2–33.7)	—	—	—	—	—	—
Moldovan (Erzya)	Uralic (Finno-Ugric)	—	—	—	46	39.1[j]	(26.4–53.6)	—	—	—	—	—	—	—	—	—
Moldovan (Moksi)	Uralic (Finno-Ugric)	—	—	—	46	21.7[j]	(12.3–35.7)	—	—	—	—	—	—	—	—	—
Moldovan (Gagauz)	Altaic (Turkic)	—	—	—	—	—	—	79	24.1[c]	(16.0–34.6)	—	—	—	—	—	—
Georgian	Caucasian (S. Caucasian)	63	14.3[f]	(7.8–25.0)	63	7.9[f]	(3.5–17.3)	—	—	—	—	—	—	—	—	—
Ossetian	IE (Indo-Iranian)	47	43.0[b]	(29.5–56.8)	47	2.0[b]	(1.3–14.3)	—	—	—	—	—	—	—	—	—
Armenian	IE (Armenian)	49	25.0[b]	(14.6–38.2)	89	6.0[b]	(3.2–13.9)	—	—	—	—	—	—	—	—	—
British	IE (Germanic)	32	68.8[i]	(51.3–82.0)	32	9.4[i]	(3.4–24.3)	—	—	—	—	—	—	—	—	—
English (Central)	IE (Germanic)	215	61.9[l]	(55.2–68.1)	215	3.3[l]	(1.6–6.6)	—	—	—	—	—	—	—	—	—
Welsh	IE (Celtic)	88	89.0[m]	(80.3–93.7)	88	1.0[m]	(0.3–6.1)	—	—	—	—	—	—	—	—	—
Orcandin	IE (Celtic)	71	66.0[m]	(54.6–76.1)	71	19.7[m]	(12.2–30.5)	—	—	—	7	0[h]	(0.0–31.2)	—	—	—
Scottish	IE (Celtic)	61	77.1[i]	(65.0–85.8)	61	6.6[i]	(2.7–15.7)	—	—	—	—	—	—	—	—	—

Table 1
Continued

Population	Language Family	N	R1b-M173[a]	95% CR	N	R1a-M17	95% CR	N	I1b* (xM26)-P37	95% CR	N	E3b1-M78	95% CR	N	J2e-M102	95% CR
Irish	IE (Celtic)	222	81.5[i]	(75.9–86.1)	222	0.5[i]	(0.1–2.5)	—	—	—	—	—	—	—	—	—
Frisian	IE (Germanic)	94	56.0[m]	(46.3–66.0)	94	7.4[m]	(3.7–14.6)	—	—	—	—	—	—	—	—	—
German	IE (Germanic)	48	47.9[f,i]	(34.4–61.7)	48	8.1[f,i]	(3.4–19.6)	16	0[c]	(0.0–16.2)	—	—	—	—	—	—
Dutch	IE (Germanic)	27	70.4[f]	(51.3–84.1)	27	3.7[f]	(0.9–18.3)	30	0[c]	(0.0–9.2)	—	—	—	34	0[g]	(0.0–8.2)
Belgian	IE (Germanic)	92	63.0[b]	(52.8–72.2)	92	4.0[b]	(1.9–10.6)	—	—	—	—	—	—	—	—	—
French	IE (Italic)	23	52.2[f]	(32.8–70.9)	23	0[f]	(0.0–11.7)	179	0[c]	(0.0–1.7)	85	4.7[h]	(1.9–11.5)	—	—	—
Bearnais	IE (Italic)	—	—	—	—	—	—	26	0[c]	(0.0–10.5)	43	4.6[g,h]	(0.9–18.3)	26	3.8[g]	(0.9–19.0)
Swiss	IE (German/Italic)	—	—	—	—	—	—	144	0[c]	(0.0–2.0)	—	—	—	—	—	—
Italian	IE (Italic)	—	—	—	332	2.7[i]	(1.4–5.1)	—	—	—	—	—	—	—	—	—
Italian (Calabria)	IE (Italic)	37	32.4[f]	(19.6–48.7)	—	—	—	148	0.7[c]	(0.2–3.7)	80	16.3[g]	(9.8–25.9)	57	1.8[g]	(0.4–9.2)
Italian (Apulia)	IE (Italic)				—	—	—	78	1.3[c]	(0.3–6.9)	86	11.6[g]	(6.5–20.1)	86	3.5[g]	(1.3–9.7)
Italian (Sardinia)	IE (Italic)	77	22.1[f]	(14.3–32.6)	—	—	—	142	0[c]	(0.0–2.1)	506	3.4[g,h]	(2.1–5.3)	144	2.1[g]	(0.8–5.9)
Italian (North-central)	IE (Italic)	50	62.0[f]	(47.9–73.5)	—	—	—	390	0.5[c]	(0.2–1.8)	212	10.4[g,h]	(7.0–15.2)	52	9.6[g]	(4.3–20.7)
Italian (South)	IE (Italic)	—	—	—	—	—	—	—	—	—	87	11.5[h]	(6.4–19.9)			
Italian (Siciliy)	IE (Italic)	—	—	—	—	—	—	—	—	—	191	13.6[g,h]	(9.5–19.2)	42	0[g]	(0.0–6.7)
Spanish	IE (Italic)	126	68.0[b]	(60.5–76.5)	126	2.0[b]	(0.5–5.6)	—	—	—	—	—	—	—	—	—
Catalan	IE (Italic)	24	79.2[f]	(59.3–90.6)	—	—	—	32	0[c]	(0.0–8.7)	33	3.0[g]	(0.7–15.3)	28	0[g]	(0.0–9.8)
Andalusian	IE (Italic)	29	65.5[f]	(47.2–80.1)	29	0[f]	(0.0–9.5)	103	0[c]	(0.0–2.8)	113	3.5[g]	(1.4–8.7)	93	1.1[g]	(0.3–5.8)
Basque (French, Spanish)	Basque (Basque)	67	88.1[f]	(78.1–93.8)	26	0[b]	(0.0–10.5)	100	0[c]	(0.0–2.9)	—	—	—	92	0	(0.0–3.2)
Portuguese	IE (Italic)	—	—	—	—	—	—	303	0.7[c]	(0.2–2.4)	—	—	—	—	—	—
Portuguese (South)	IE (Italic)	57	56.0[b]	(43.2–68.3)	57	2.0[b]	(0.4–9.2)	—	—	—	—	—	—	—	—	—
Portuguese (North)	IE (Italic)	328	62.0[b]	(56.5–67.0)	328	0[b]	(0.0–0.9)	—	—	—	—	—	—	—	—	—
Turkish	Altaic (Turkic)	523	16.3[n]	(13.3–19.7)	523	6.9[n]	(5.2–9.6)	741	2.3[c]	(1.4–3.6)	—	—	—	—	—	—
Anatolian (Central)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	61	6.6[h]	(2.7–15.7)	—	—	—
Turkish (Istanbul)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	81	8.6[g,h]	(4.3–16.8)	73	0[g]	(0.0–4.0)
Turkish (Konya)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	117	12.8[g]	(8.0–20.1)	129	0.8[g]	(0.2–4.2)
Turkish (Cypriot)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	46	13.0[h]	(6.2–25.7)	—	—	—
Turkish (Southeastern)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	24	4.2[h]	(1.0–20.4)	—	—	—
Turkish (Erzurum)	Altaic (Turkic)	—	—	—	—	—	—	—	—	—	25	4.0[h]	(0.9–19.6)	—	—	—
Algerian	Afro-Asiatic (Semitic)	35	0[o]	(0.0–8.0)	35	0[o]	(0.0–8.0)	35	0[o]	(0.0–8.0)	32	6.3[g]	(1.9–20.2)	55	0[g,o]	(0.0–5.2)
Tunisian	Afro-Asiatic (Semitic)	148	6.7[o]	(3.7–12.0)	148	0[o]	(0.0–2.0)	148	0[o]	(0.0–2.0)	58	15.5[g]	(8.4–27.0)	73	1.4[g]	(0.3–7.3)
Moroccan	Afro-Asiatic (Semitic)	44	6.8[o]	(2.5–18.3)	44	0[o]	(0.0–6.4)	44	0[o]	(0.0–6.4)	93	27.9[g]	(19.9–37.8)	44	0[o]	(0.0–6.4)

NOTE.—IE = Indo-European.

[a] R1b haplogroup is defined by M173 (xSRY-1532); data from Rosser et al. (2000), Helgason et al. (2000), Weale et al. (2002), and Wilson et al. (2001) are deducted based on the last typed marker 92R7 (xSRY-1532).

[b] Rosser et al. (2000).

[c] Rootsi et al. (2004).

[d] This study.

[e] Barać et al. (2003).

[f] Semino et al. (2000).

[g] Semino et al. (2004).

[h] Cruciani et al. (2004).

[i] Helgason et al. (2000).

[j] Malaspina et al. (2003).

[k] Tambets et al. (2004).

[l] Weale et al. (2002).

[m] Wilson et al. (2001).
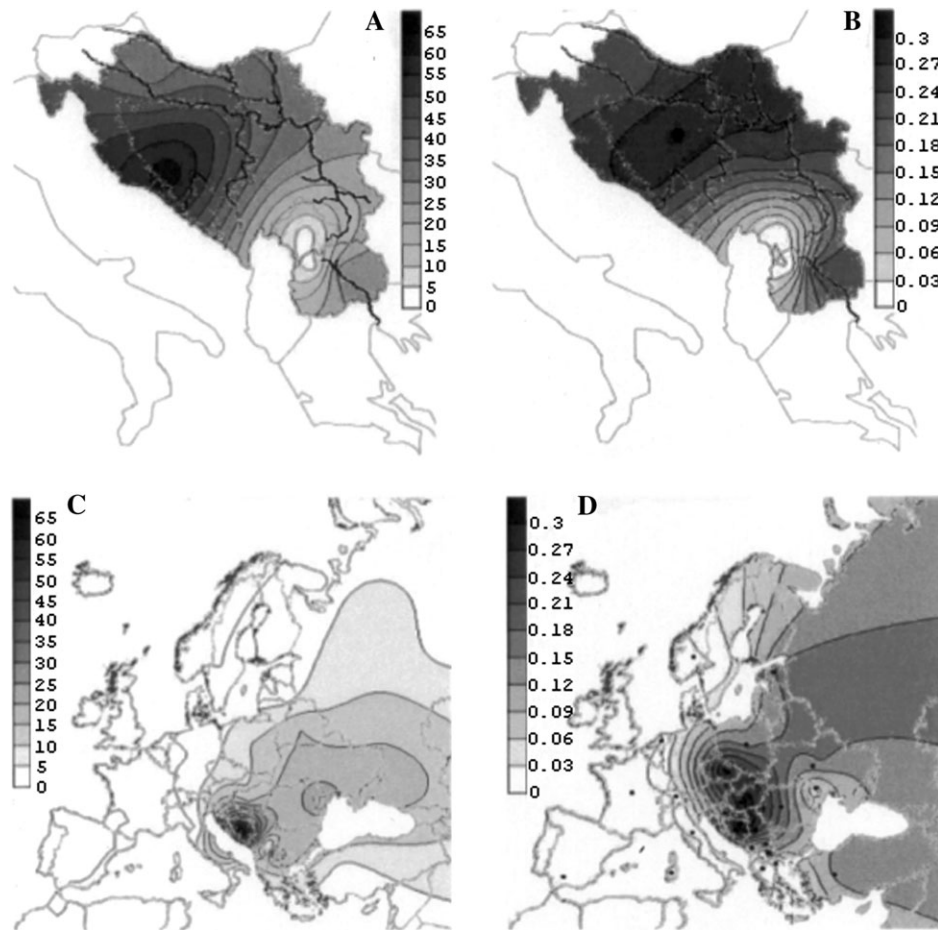
[n] Cinnioğlu et al. (2004)

[o] Arredi et al. (2004).

FIG. 3.—I1b* (xM26) frequency and variance surfaces in SEE (panels *A* and *B*) were generated from the data in this study. I1b* (xM26) frequency surfaces in Europe, northern Africa, and Asia Minor (panel *C*) were generated from the data reported in table 1, and variance surfaces (panel *D*) were generated from STR data in this study and Rootsi et al. (2004).

out of which the α cluster, largely characterized by an A7.1 nine-repeat allele, is confined to Europe (the Balkans) and Turkey (Cruciani et al. 2004). E3b1 variance distribution depicted in figure 4(*D*) does not overlap with its frequency distribution possibly because analyzed E3b1 chromosomes harbor diverse background motifs. It is very likely that a variance peak centered in northeastern Africa as well as high variance values in Turkey and southern Italy are due to the inclusion of δ (and a few southern Italian β) chromosomes. Almost 93% of SEE E3b1 chromosomes are classified into α cluster. In Europe, the highest E3b1α variance is among Apulians, Greeks, and Macedonians, and the highest frequency of the cluster is among Albanians, Macedonians, and Greeks (table 1). Bearing in mind the congruent E3b1α frequency, variance maximums, and star-like phylogenetic network (fig. 8*B*), it is possible to envision that a yet undefined sublineage downstream of M78, characterized by the nine-repeat allele at A7.1 locus, may have originated in south Europe and SEE from where it dispersed in different directions. Furthermore, it may be envisioned that the observed E3b1α frequency distribution in Anatolia might stem from a back migration originating in south Europe and SEE. Our estimated range expansion of 7.3 ± 2.8 KYA is close to the 7.8 KYA (95% CI 6.3–9.2 KYA) estimate for expansions of cluster α chromo-

somes in Europe reported by Cruciani et al. (2004) and the 6.4 KYA estimate for E3b1-M78 STR variance in Anatolia dated by Cinnioğlu et al. (2004). The frequency and variance decline of E3b1 in SEE is rather continuous (fig. 4*A* and *B*), with a frequency peak extending from the southeastern edge of the region and a variance peak in southwest. Observed high E3b1 frequency in Kosovar Albanians (46%) and Macedonian Romani (30%) represent a focal rather than a clinal phenomenon resulting most likely from genetic drift. E3b1 frequency and variance are significantly correlated with latitude, showing higher values toward south (table 2), both when all SEE populations are considered ($r = -0.51$, $P = 0.05$, for frequency and $r = -0.706$, $P = 0.05$, for variance) and when Kosovar Albanians and Macedonian Romani are excluded ($r = -0.597$, $P = 0.05$, for frequency and $r = -0.676$, $P = 0.05$, for variance). A lower frequency of E3b1 significantly distinguishes populations of the Adriatic-Dinaric complex, i.e., mainland Croatians, Bosnians, and Herzegovinians (7.9%; 95% CI 0.054–0.114), from their neighboring populations of the Vardar-Morava-Danube river system, i.e., Serbians and Macedonians (21.9%; 95% CI 0.166–0.283). These observations hint a mosaic of different E3b1 dispersal modes over a short geographic distance and point to the Vardar-Morava-Danube river system
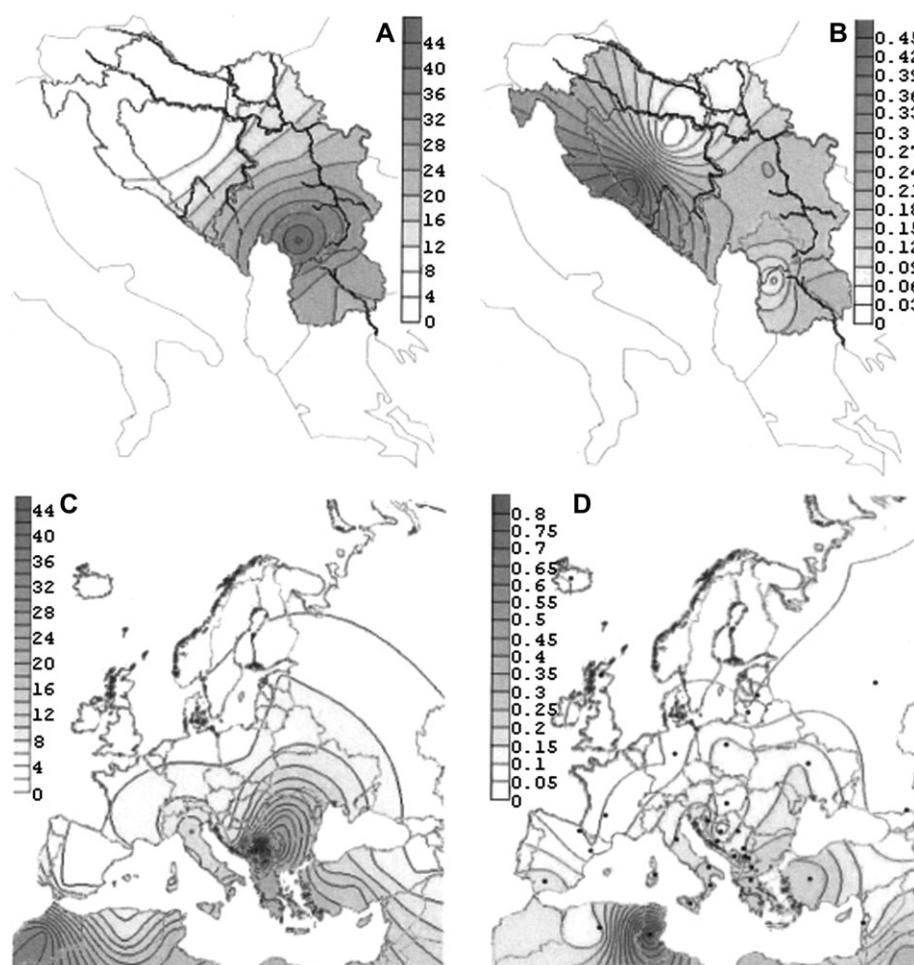
Fig. 4.—E3b1 frequency and variance surfaces in SEE (panels *A* and *B*) were generated from the data in this study. E3b1 frequency surfaces in Europe, northern Africa, and Asia Minor (panel *C*) were generated from the data reported in table 1, and variance surfaces (panel *D*) were calculated from STR data in this study and Semino et al. (2004).

as one of major routes for E3b1, in fact E3b1α, expansion from south and southeastern to continental Europe. In fact, dispersals of farmers throughout the Vardar-Morava-Danube catchments basin are also evidenced in the archaeological record (Tringham 2000).

R1a haplogroup occurs at 16% frequency in SEE (fig. 2). The age of M17 has been approximated to 15 KYA (Semino et al. 2000; Wells et al. 2001). Kivisild et al. (2003) suggested that southern and western Asia might be the source of R1 and R1a differentiation. Current R1a-M17/SRY-1532 distribution in Europe shows an increasing west-east frequency and variance gradients with peaks among Finno-Ugric and Slavic speakers (fig. 5*C* and *D*). Similar to I1b* (xM26), R1a frequency gradient decreases slowly to the south (to 10% in Albanians, 8% in Greeks, and 7% in Turks) and abruptly in the west (3% in Italians) (table 1). R1a frequency and STR variance decrease in the north-south direction in SEE, from 34%–25% in mainland Croatians and Bosnians to 12%–16% in Herzegovinians, Macedonians, and Serbians (fig. 5*A* and *B*). Moreover, R1a frequency is significantly correlated with latitude (table 2) when all studied SEE populations are considered ($r = 0.865$, $P = 0.01$) and also when Kosovar

Albanians and Macedonian Romani are excluded ($r = 0.743$, $P = 0.01$). High R1a haplotype diversity in SEE is evident in the phylogenetic network (fig. 8*C*) and the estimated range expansion at 15.8 ± 2.1 KYA, consistent with its deep Paleolithic time depth, as previously suggested (Semino et al. 2000; Wells et al. 2001). At this level of resolution, it is not clear what temporal and effective population size differences contributed to this deep Paleolithic signal as high R1a variance in SEE might be explained by either ancient demography or more recent bottlenecks and founder effects in different Slavic tribes. At least three major episodes of gene flow might have enhanced R1a variance in the region: early post-LGM recolonizations expanding from the refugium in Ukraine, migrations from northern Pontic steppe between 3000 and 1000 B.C., as well as possibly massive Slavic migration from A.D. 5th to 7th centuries.

R1b haplogroup is present in SEE at a level of 9% (fig. 2). R1b-M173 lineages are considered to trace an Upper Paleolithic migration from West Asia to European regions then occupied by Aurignacian culture (Semino et al. 2000; Underhill et al. 2001; Wells et al. 2001). The spatial distribution of R1b lineages shows a frequency peak
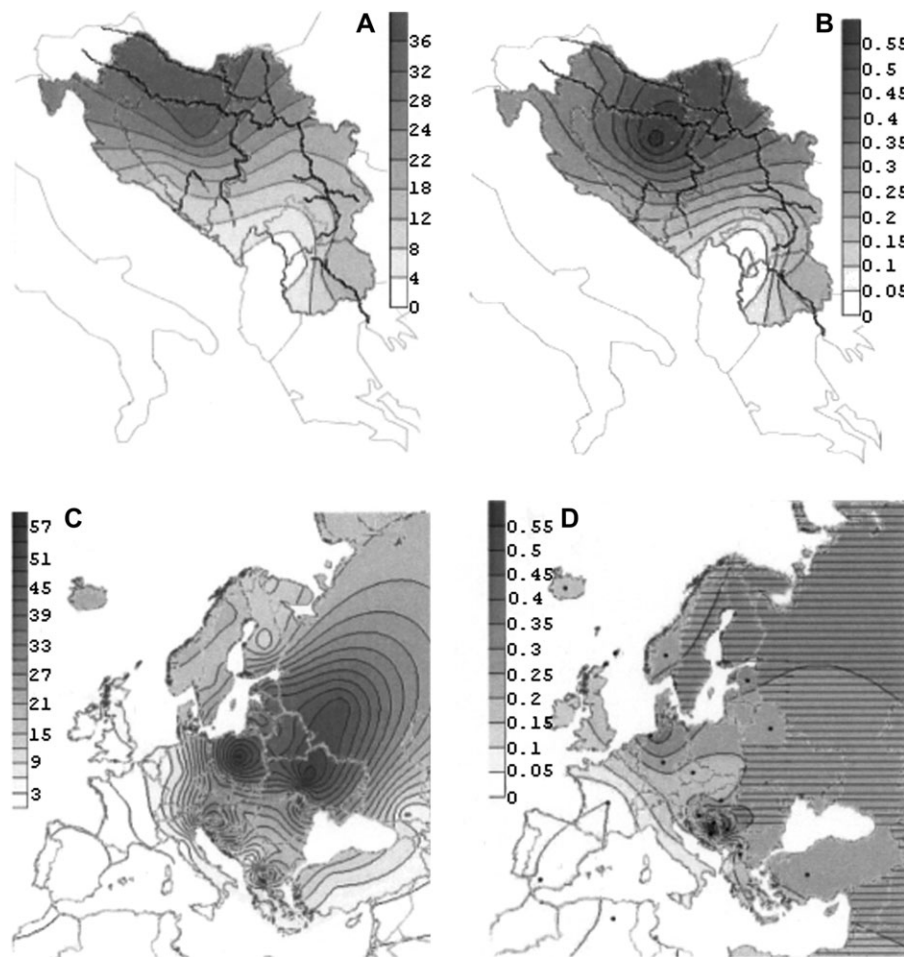
FIG. 5.—R1a frequency and variance surfaces in SEE (panels *A* and *B*) were generated from the data in this study. R1a frequency surfaces in Europe, northern Africa, and Asia Minor (panel *C*) were generated from the data reported in table 1, and variance surfaces (panel D) were calculated from STR data in this study, Rootsi et al. unpublished data, Cinnioğlu et al. (2004), Behar et al. (2003), Weale et al. (2002), Wilson et al. (2001), Helgason et al. (2000), and Hurles et al. (1999). Shaded areas in panel *D* correspond to regions for which combined SNP and STR Y chromosomal data are not available.

(40%–80%) in western Europe and a decrease in eastern (with the exception of 43% in the Ossetians) and southern Europe (fig. 6*C*), whereas R1b variance shows multiple peaks in West Europe and Asia Minor (fig. 6*D*). While R1b variance displays a clear-cut northwestern-southeastern decline in SEE (fig. 6*B*), R1b frequency decline continues from western toward southeastern and southern Europe, but two intermediate local peaks are evident, in north among mainland Croatians and Serbians and in south among Kosovar Albanians, Albanians, and Greeks (fig. 6*C*). These spatial patterns might be due to the fact that R1b lineages contain associated RFLP 49a,f ht 15 and 35 sublineages with opposite distributions possibly reflecting repeopling of Europe from Iberia and Asia Minor during the Late Upper Paleolithic and Holocene (Cinnioğlu et al. 2004). The overall R1b frequency distribution in the Balkan Peninsula suggests its possible arrival from two different source populations during recolonization of Europe. We estimated the range expansion of R1b lineages in SEE at 11.6 ± 1.4 KYA. Although R1b lineages could have accumulated STR variance before diffusion in SEE, it is significant that its estimated range expansion almost per-

fectly matches the coalescent estimate for the I1b* (xM26) lineages, pointing to the YD to Holocene transition as possibly a period when these two major Y chromosome lineages started to expand in the region.

Haplogroup J defined by a 12f2 polymorphism is subdivided into two major clades, J1-M267 and J2-M172 (Cinnioğlu et al. 2004). J2-M172 is more prevalent in Europe where at least five different lineages can be traced—J2e*-M102, J2e1-M241, J2*-M172, J2f*-M67, and J2f1-M92 (fig. 2, Semino et al. 2004). In SEE, the most frequent are J2e lineages that comprise 5% of all chromosomes, while J2f cluster, a predominant J2 cluster in Greeks and Italians (Di Giacomo et al. 2004), is present at a frequency less than 1% (fig. 2). Most likely due to genetic drift, Kosovar Albanians harbor a J2e frequency peak whereas variance maximum declines from the southeastern edge of the studied region (fig. 7*A* and *B*). Even though J2e frequencies do not correlate with geography, J2e variances show significant correlations with latitude and longitude and are highest toward south and east of the region (table 2). The correlation between geography and haplogroup frequencies are significant when all SEE
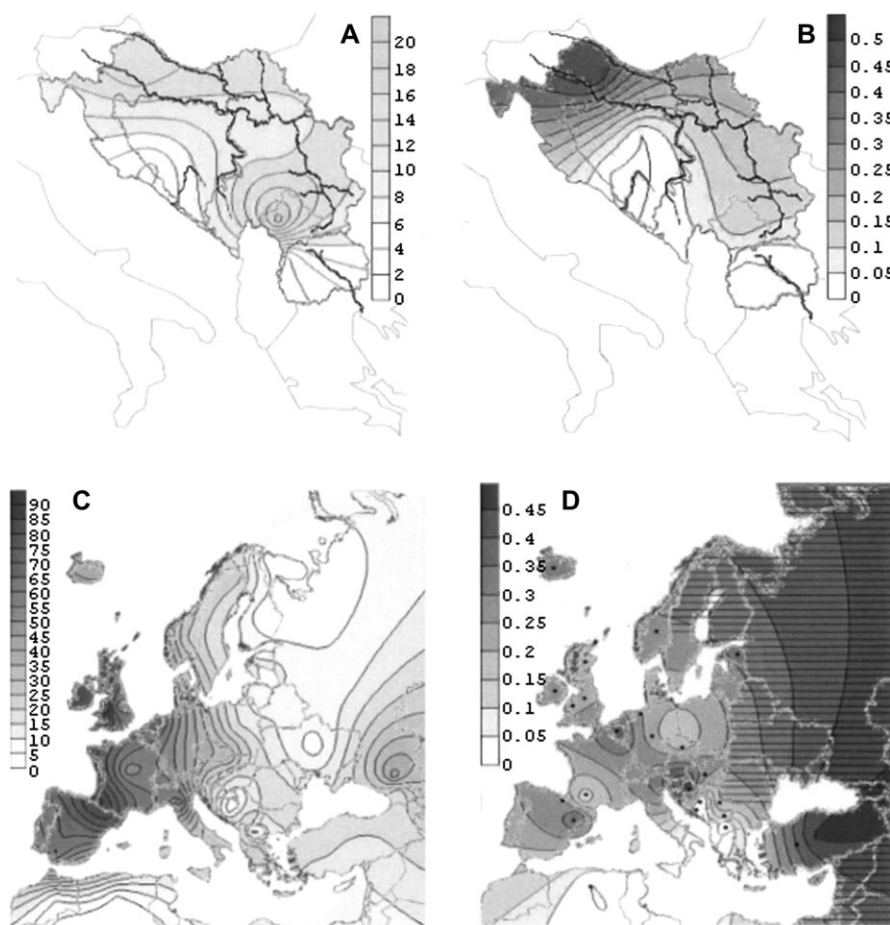
Fig. 6.—R1b frequency and variance surfaces in SEE (panels *A* and *B*) were generated from the data in this study. R1b frequency surfaces in Europe, northern Africa, and Asia Minor (panel *C*) were generated from the data reported in table 1, and variance surfaces (panel *D*) were calculated from STR data in this study, Rootsi et al. unpublished data, Cinnioğlu et al. (2004), Behar et al. (2003), Weale et al. (2002), Wilson et al. (2001), Helgason et al. (2000), and Hurles et al. (1999). Shaded areas in panel *D* correspond to regions for which combined SNP and STR Y chromosomal data are not available.

populations are considered ($r = -0.949$, $P = 0.05$) and when Kosovar Albanians and Macedonian Romani are excluded ($r = -0.949$, $P = 0.05$). Our estimated range expansion for J2e at 2.8 ±1.6 KYA (for all SEE populations) and 3 ± 1.9 KYA (SEE populations without Kosovar Albanians) succeeds the dates of 7.9 ± 2.3 KYA (Semino et al. 2004) and 8.6 KYA (Cinnioğlu et al. 2004). The J2e-M102 spatial distribution depicted in figure 7(*C* and *D*) with two frequency and variance peaks positioned in the Balkans and central Italy may be explained by the maritime spread of J2e lineages from southern Balkans toward Apennines at times later than those based on the classical model of demic expansions carried by Neolithic agriculturists from the Middle East via Balkans toward rest of Europe.

Widely spread Romani haplogroup H1 is a major lineage cluster in Macedonian Romani (fig. 2). A 2-bp deletion at M82 locus defining this haplogroup was also reported in one-third of males from traditional Romani populations living in Bulgaria, Spain, and Lithuania (Gresham et al. 2001). Its ancestral M52 A → C transversion was reported in the Vlax Roma (Kalaydjieva et al. 2001) and India (Ramana et al. 2001; Wells et al. 2001; Kivisild et al. 2003). Out of 34 H1-M82 males, 10 were typed for mito-

chondrial DNA (mtDNA) and belonged to haplogroup M that was highly frequent in Macedonian Romani (Cvjetan et al. 2004), traditional Romani populations (Gresham et al. 2001), and India (Kivisild et al. 2003). High prevalence of Asian-specific Y chromosome haplogroup H1 and mtDNA haplogroup M supports their Asian (Indian) origin and a hypothesis of a small number of founders diverging from a single ethnic group in India (Gresham et al. 2001).

F*, G-M201, K* (xP), P* (xR1, Q), and Q-M242 lineages occur at low frequencies in SEE (fig. 2). The Herzegovinian Q-M242 sample harbors a STR motif previously seen in eastern Adriatic haplogroup Q lineages that are marked by the typical presence of the unusually long DYS392-15 allele (Barać et al. 2003).

We conclude that even though the majority of identified SEE paternal lineages are consistent with the typical European Y chromosome gene pool, their distribution and estimated range expansions clarify the specific role of this region in structuring the European genetic landscape. Contemporary Slavic paternal gene pool is characterized by the predominance of R1a and I1b* (xM26) variants as well as the scarcity of E3b1 lineages as a result of the following prehistoric gene flows. First, we envision the post-LGM
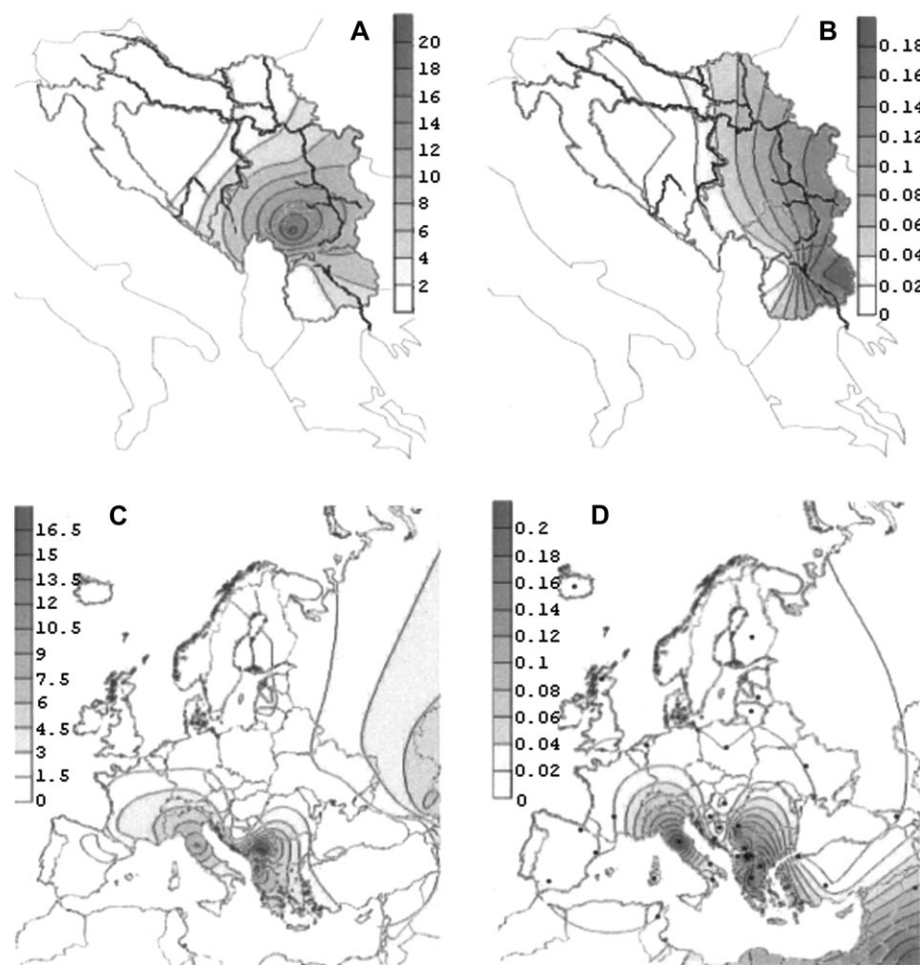
FIG. 7.—J2e frequency and variance surfaces in SEE (panels *A* and *B*) were generated from the data in this study. J2e frequency surfaces in Europe, northern Africa, and Asia Minor (panel *C*) were generated from the data reported in table 1, and variance surfaces (panel *D*) were generated from STR data in this study and Semino et al. (2004).

R1a expansion from eastern to western Europe and second the YD-Holocene I1b* (xM26) diffusion out of the Balkans in addition to subsequent R1a and I1b* (xM26) putative gene flows between eastern Europe and SEE. Lastly, we envision a weaker extent of E3b1 dispersal out of southern Europe and SEE toward eastern Europe rather than toward western (especially Mediterranean) Europe. Our results also stress that I1b* (xM26) wide geographic distribution and

Table 2
Correlations of Major Y Chromosome Haplogroup Frequencies and Variances with Geography

Haplogroup	*n*	Spearman's *r* (Frequency)		Spearman's *r* (Variance)	
		Latitude	Longitude	Latitude	Longitude
I1b* (xM26)	207	−0.077	−0.084	−0.385	0.280
I1b* (xM26) (without KA and MR)[a]	202	−0.464	0.327	−0.455	0.345
I1a	28	−0.491	0.249	−0.374	0.244
I1a (without KA and MR)	21	−0.584	0.299	−0.457	0.203
R1a	104	0.865*	−0.480	0.385	0.238
R1a (without KA and MR)	98	0.743*	−0.159	0.491	0.191
R1b	63	0.175	−0.238	0.118	−0.209
R1b (without KA and MR)	38	0.364	−0.373	0.103	−0.079
E3b1	131	−0.651**	0.501	−0.706**	0.572
E3b1 (without KA and MR)	62	−0.597**	0.360	−0.676**	0.511
J2e1	26	−0.148	0.304	−0.949**	0.949**
J2e1 (without KA and MR)	6	0.042	0.806	−0.949**	0.949**

[a] Population abbreviations: KA = Kosovar Albanians, MR = Macedonian Romani.
* Correlation is significant at the 0.01 level (two-tailed).
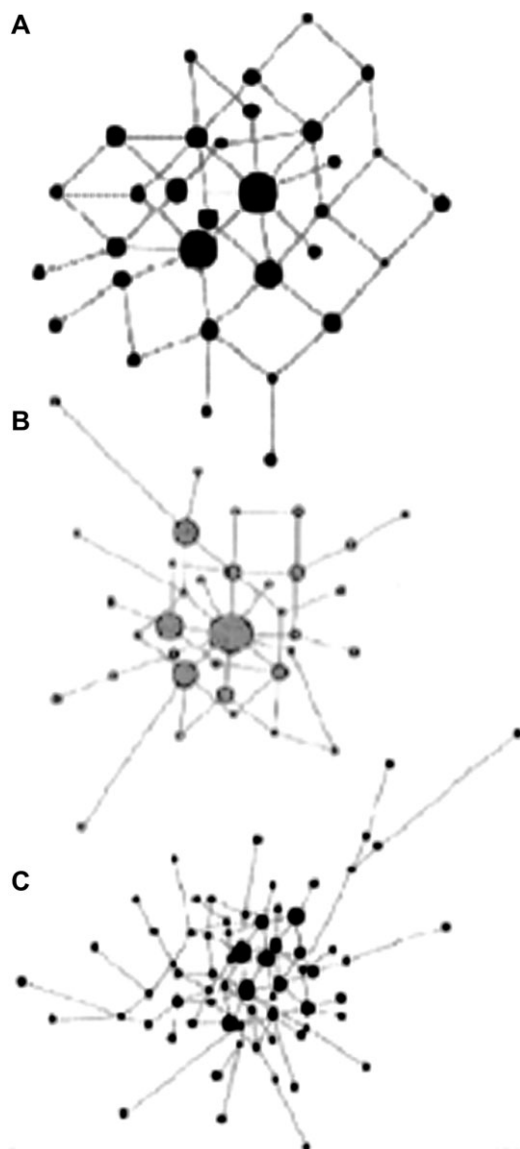** Correlation is significant at the 0.05 level (two-tailed).

FIG. 8.—Microsatellite networks of major Y chromosomal lineages in SEE: (*A*) I1b* (xM26) (*B*) E3b1α; (*C*) R1a. Microsatellite haplotypes are represented by circles, with areas proportional to the number of individuals harboring the haplotype. Smallest circle represents single haplotype in panel *B* and *C* and two haplotypes in panel *A*. Branch lengths are proportional to the number of one-step mutations separating two haplotypes.

massive frequencies accompanied with high diversity in most of its range among major SEE populations testify impressively to their common paternal history, whereas observed genetic heterogeneity structured mostly along the northwestern-southeastern axis is a result of attested prehistoric and historical gene flows with different temporal and directional characteristics. Yet the main difference between the paternal genetic history of the Slavic-speaking populations lies in the presence, among eastern Slavs (Russians, Ukrainians, Belarussians), of haplogroup N chromosomes, virtually absent among any of the western or southern Slavic populations (Rosser et al. 2000; Semino et al. 2000; Barać et al. 2003; Tambets et al. 2004), unequivocally suggesting that the historic eastward expansion of Slavs in the middle of the first millennium A.D. resulted

in a substantial admixture of them with the substratum populations, inhabiting East Europe, among whom this largely northern Eurasian haplogroup was and still is widely spread.

Literature Cited

Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, N.J.

Arredi, B., E. S. Poloni, S. Paracchini, T. Zerjal, D. M. Fathallah, M. Makrelouf, V. L. Pascali, A. Novelletto, and C. Tyler-Smith. 2004. A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. Am. J. Hum. Genet. **75**:338–345.

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

Barać, L., M. Peričić, I. Martinović Klarić, S. Rootsi, B. Janićijević, T. Kivisild, J. Parik, I. Rudan, R. Villems, and P. Rudan. 2003. Y chromosomal heritage of Croatian population and its island isolates. Eur. J. Hum. Genet. **11**:535–542.

Behar, D. M., M. G. Thomas, K. Skorecki et al. (12 co-authors). 2003. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am. J. Hum. Genet. **73**:768–779.

Childe, V. G. 1958. The prehistory of European society. Penguin Books, London.

Cinnioğlu, C., R. King, T. Kivisild et al. (15 co-authors). 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum. Genet. **114**:127–148.

Cruciani, F., R. La Fratta, P. Santolamazza et al. (19 co-authors). 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am. J. Hum. Genet. **74**:1014–1022.

Cruciani, F., P. Santolamazza, P. D. Shen et al. (16 co-authors). 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am. J. Hum. Genet. **70**:1197–1214.

Cvjetan, S., H.-V. Tolk, L. Barać Lauc et al. (15 co-authors). 2004. Frequencies of mtDNA haplogroups in southeastern Europe—Croatians, Bosnians and Herzegovinians, Serbians, Macedonians and Macedonian Romani. Coll. Antropol. **28**:193–198.

Di Giacomo, F., F. Luca, L. O. Popa et al. (27 co-authors). 2004. Y chromosomal haplogroup J as a signature of the post-Neolithic colonization of Europe. Hum. Genet. **115**:357–371.

Gresham, D., B. Morar, P. A. Underhill et al. (17 co-authors). 2001. Origins and divergence of the Roma (gypsies). Am. J. Hum. Genet. **69**:1314–1331.

Hammer, M. F., and S. Horai. 1995. Y chromosomal DNA variation and the peopling of Japan. Am. J. Hum. Genet. **56**: 951–962.

Helgason, A., S. Siguroardottir, J. Nicholson, B. Sykes, E. W. Hill, D. G. Bradley, V. Bosnes, J. R. Gulcher, R. Ward, and K. Stefansson. 2000. Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am. J. Hum. Genet. **67**:697–717.

Hurles, M. E., R. Veitia, E. Arroyo et al. (18 co-authors). 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am. J. Hum. Genet. **65**:1437–1448.

Kaessmann, H., F. Heissig, A. von Haeseler, and S. Pääbo. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. **22**:78–81.

Kalaydjieva, L., F. Calafell, M. A. Jobling et al. (11 co-authors). 2001. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur. J. Hum. Genet. **9**:97–104.

Kayser, M., A. Caglia, D. Corach et al. (26 co-authors). 1997. Evaluation of Y-chromosomal STRs—a multicenter study. Int. J. Legal Med. **110**:125–133.

Kivisild, T., S. Rootsi, M. Metspalu et al. (18 co-authors). 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. **72**: 313–332.

Malaspina, P., A. I. Kozlov, F. Cruciani et al. (11 co-authors). 2003. Analysis of Y-chromosome variation in modern populations at the European-Asian border. Pp. 309–313 *in* K. Boyle, C. Renfrew, and M. Levine, eds. Ancient interactions: east and west in Eurasia. McDonald Institute for Archaeological Research Monograph Series, Cambridge University Press, Cambridge.

Mathias, N., M. Bayés, and C. Tyler-Smith. 1994. Highly informative compound haplotypes for the human Y chromosomes. Hum. Mol. Genet. **3**:115–123.

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1998. A simple salting our procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. **18**: 2125.

Ramana, G. V., B. Su, L. Jin, L. Singh, N. Wang, P. Underhill, and R. Chakraborty. 2001. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. Eur. J. Hum. Genet. **9**:695–700.

Renfrew, A. C. 1987. Archaeology and language: the puzzle of Indo-European origins. Cape, London.

Richards, M., V. Macaulay, E. Hickey et al. (26 co-authors). 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am. J. Hum. Genet. **67**:1251–1276.

Rootsi, S., C. Magri, T. Kivisild et al. (45 co-authors). 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am. J. Hum. Genet. **75**:128–137.

Rosser, Z. H., T. Zerjal, M. E. Hurles et al. (63 co-authors). 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. **67**:1526–1543.

Semino, O., C. Magri, G. Benuzzi et al. (16 co-authors). 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am. J. Hum. Genet. **74**:1023–1034.

Semino, O., G. Passarino, P. J. Oefner et al. (17 co-authors). 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science **290**:1155–1159.

Tambets, K., S. Rootsi, T. Kivisild et al. (46 co-authors). 2004. The western and eastern roots of the Saami—the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am. J. Hum. Genet. **74**:661–682.

Tringham, R. 2000. Southeastern Europe in the transition to agriculture in Europe: bridge, buffer or mosaic. Pp. 19–56 *in* T. D. Price, ed. Europe's first farmers, Cambridge University Press, Cambridge.

Underhill, P. A., G. Passarino, A. A. Lin, P. Shen, M. M. Lahr, R. A. Foley, P. J. Oefner, and L. L. Cavalli-Sforza. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. **65**:43–62.

Underhill, P. A., P. D. Shen, A. A. Lin et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. Nat. Genet. **26**:358–361.

Weale, M. E., D. A. Weiss, R. F. Jager, N. Bradman, and M. G. Thomas. 2002. Y chromosome evidence for Anglo-Saxon mass migration. Mol. Biol. Evol. **19**:1008–1021.

Wells, R. S., N. Yuldasheva, R. Ruzibakiev et al. (27 co-authors). 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc. Natl. Acad. Sci. USA **98**: 10244–10249.

White, P. S., O. L. Tatum, L. L. Deaven, and J. L. Longmire. 1999. New, male-specific microsatellite markers from the Y chromosome. Genomics **57**:433–437.

Whitfield, L. S., J. E. Sulston, and P. N. Goodfellow. 1995. Sequence variation of the human Y chromosome. Nature **378**:379–380.

Wilson, J. F., D. A. Weiss, M. Richards, M. G. Thomas, N. Bradman, and D. B. Goldstein. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc. Natl. Acad. Sci. USA **98**:5078–5083.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. **12**:339–348.

Zhivotovsky, L. A., P. A. Underhill, C. Cinnioglu 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.

Zvelebil, M., and M. Lillie. 2000. Transition to agriculture in eastern Europe. Pp. 57–92 *in* T. D. Price, ed. Europe's first farmers. Cambridge University Press, Cambridge.