Understanding Language Evolution Using an Event-Based Model

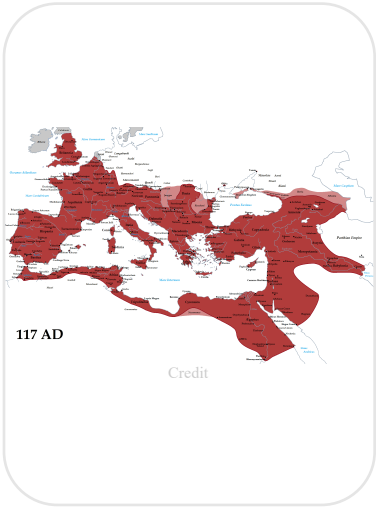
John Huelsenbeck
Department of Integrative Biology
johnh@berkeley.edu

Introduction

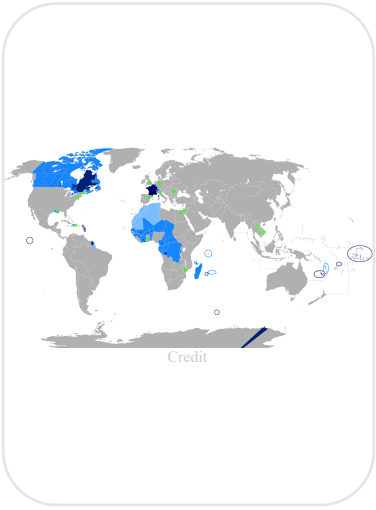
In this paper, we attempt to do the impossible!

Languages

Latin

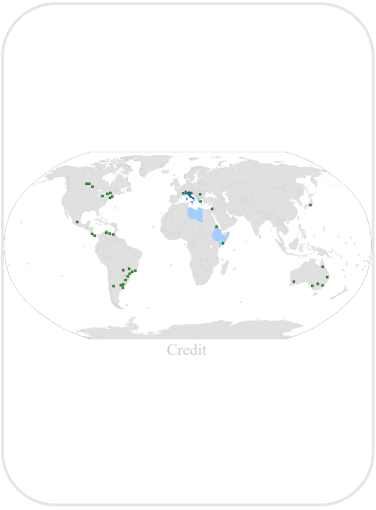


French



Spanish

Italian

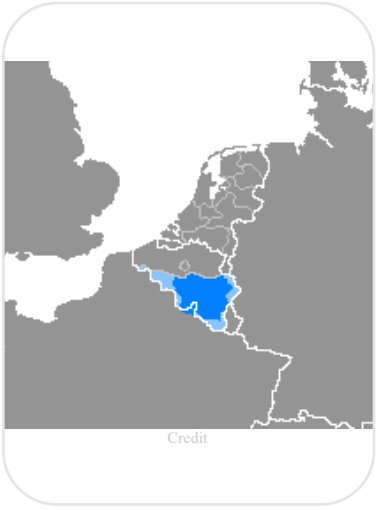


Brazilian Portuguese

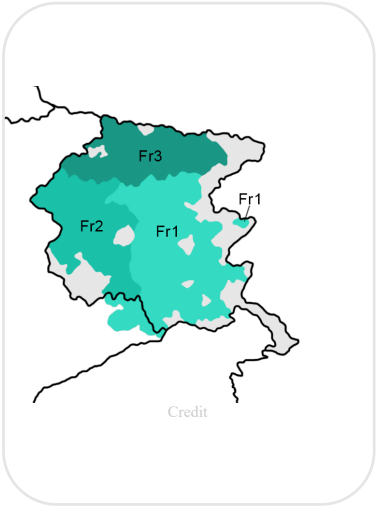
Portuguese

Catalan

Walloon



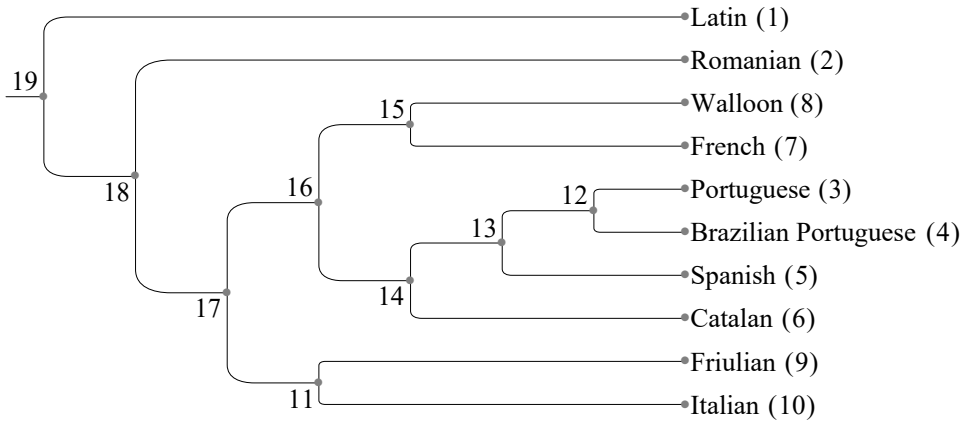
Friulian



Romanian

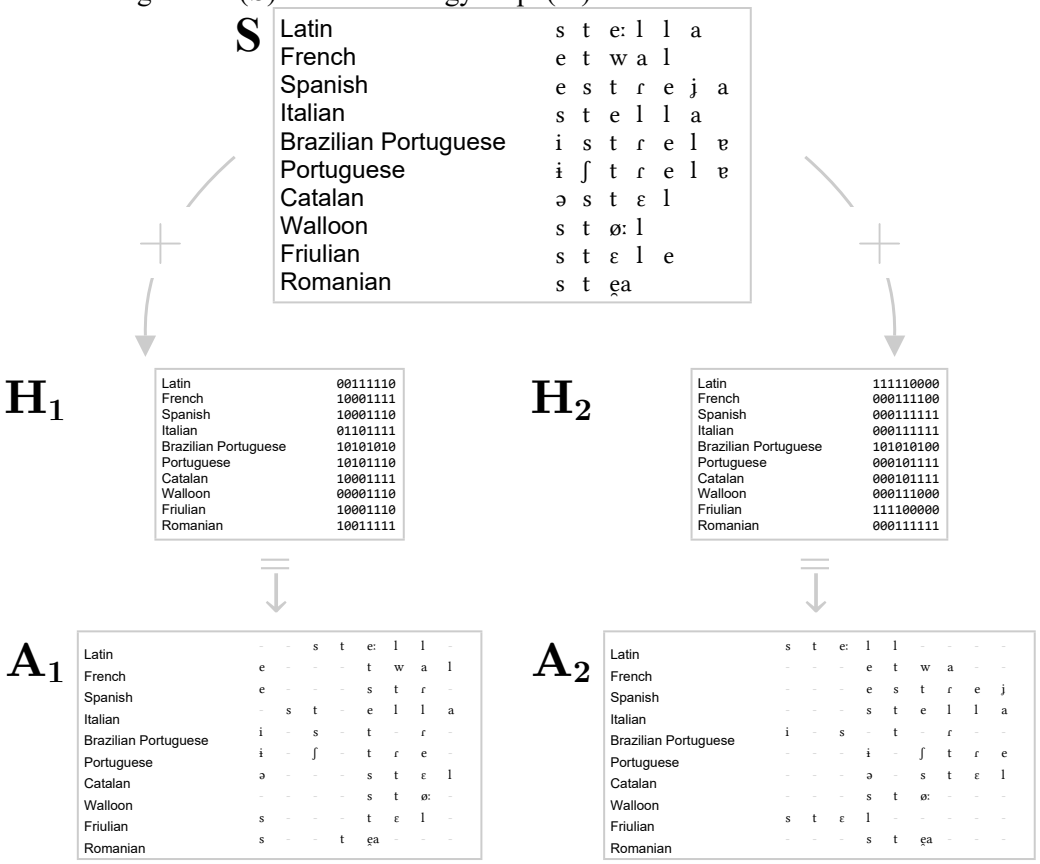
Example Tree

An example tree showing the relationships of $N = 10$ languages.



Alignment

Alignments (**A**) are formed from the observed segments (**S**) and a homology map (**H**).



Character Assignments

Each segment gets a different number

1	e	2	g	3	oː	4	ʒ	5	ə	6	ʝ	7	o	8	i	9	w	10	ɔ	11	ʤ	12	ɪ	13	j	14	n	15	s	16	u	17	ʃ
18	oj	19	v	20	b	21	t	22	uː	23	y	24	d	25	f	26	eː	27	m	28	a	29	ɾ	30	ɐ	31	œ	32	ɛ				
33	l	34	r	35	x	36	ʎ	37	ʁ	38	iː	39	ɲ	40	o̯a	41	h	42	ẽj̃	43	ɔ̃	44	ɑ̃	45	ĩ	46	ɐ̃	47	ʧ				
48	ɨ	49	ã	50	p	51	z	52	ð	53	k	54	ʦ	55	aː	56	ɐ̃j̃	57	ʀ	58	ɒ	59	θ	60	ej	61	ɐj	62	õ				
63	ɛː	64	ɥ	65	øː	66	ɬ	67	c	68	ɛj	69	ø	70	e̯a	71	k^{w}	72	ɣ	73	aj	74	g	75	ẽ	76	ŋ	77	g^{w}				
78	ẽu̯	79	ɑ	80	β	81	ɟ	82	ɛ̃	83	ʧ̃	84	ɻ	85	ɔː	86	ũ	87	j̃	88	œ̃	89	au̯	90	w̃								

Partition Assignments

Basic Rules

1 **Nasal Vowel**

ẽj̃ ɔ̃ ɑ̃ ĩ ɐ̃ ã ɐ̃j̃ õ ẽ ẽu̯ ɛ̃ ũ œ̃

2 **Vowel**

e oː ə o i ɔ ɪ u oj uː y eː a ɐ œ ɛ iː o̯a ɨ aː ɒ ej ɐj ɛː øː ɛj ø e̯a aj ɑ ɔː au̯

3 **Nasal Consonant**

n m ɲ ŋ j̃ w̃

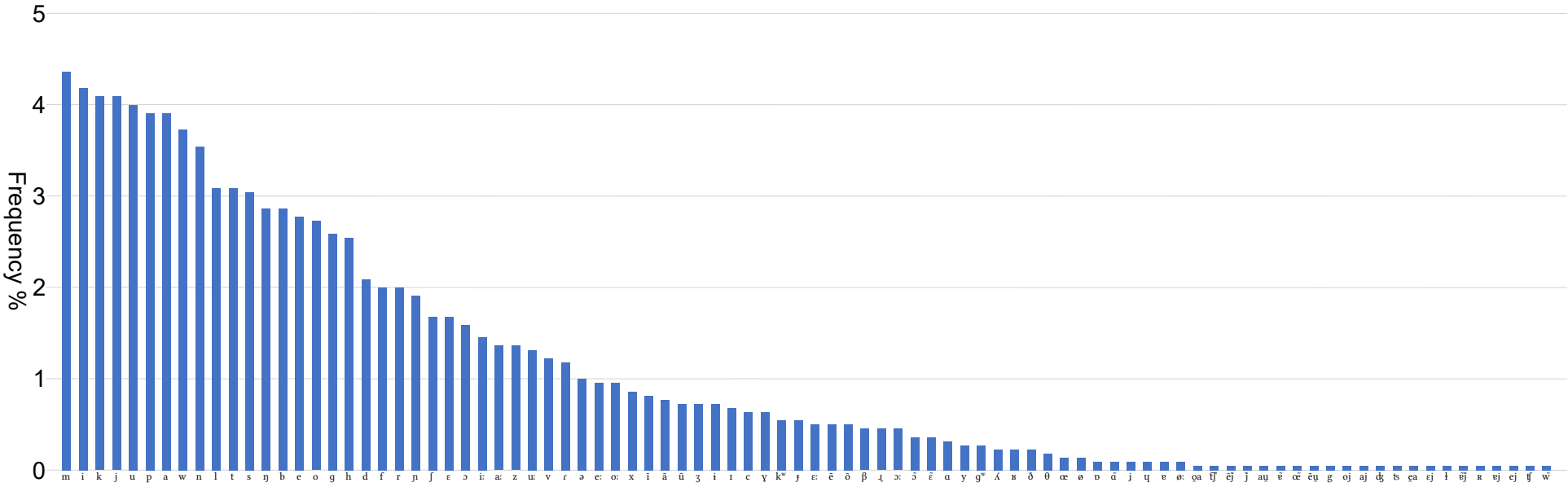
4 **Non Sylabic Sonorant**

w j l r

5 **Consonant**

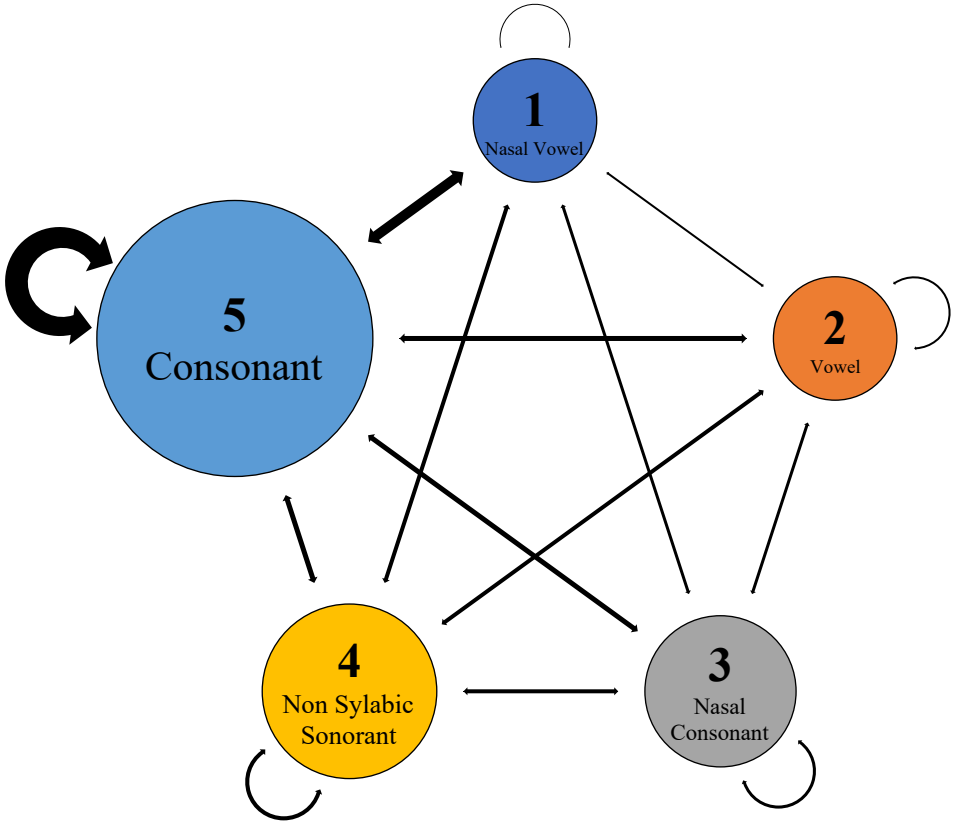
g ʒ ʝ ʤ s ʃ v b t d f ɾ x ʎ ʁ h ʧ p z ð k ʦ ʀ θ ɥ ɬ c kʷ ɣ g gʷ β ɟ ʧ ɻ

Prior Segment Frequencies

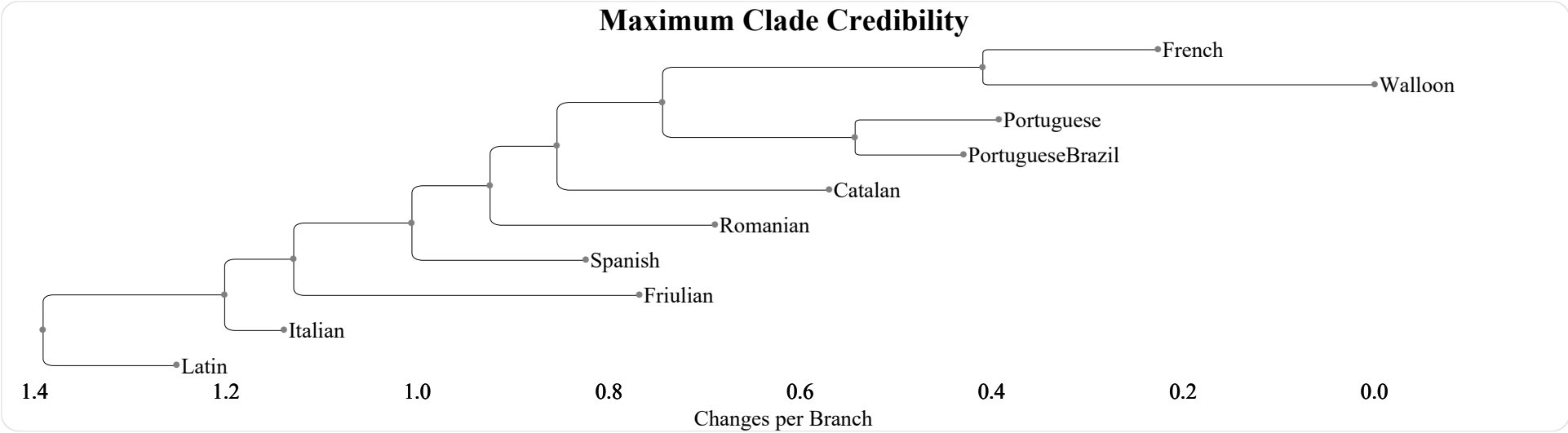


Transition Rates Between Groups

For the 'Basic Rules' model, states were grouped into five sets: Nasal Vowel (1), Vowel (2), Nasal Consonant (3), Non Sylabic Sonorant (4) and Consonant (5). Here, the area of the circles is proportional to the estimated equilibrium frequencies for each group. The width of the arrows is proportional to the estimated rates. Note that rates are higher for transitions from one word segment to another when the word segments are in the same group.



Results


Maximum Clade Credibility
French
Walloon
Portuguese
PortugueseBrazil
Catalan
Romanian
Spanish
Friulian
Italian
Latin
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Changes per Branch

Questions