

Understanding Language Evolution Using an Event-Based Model

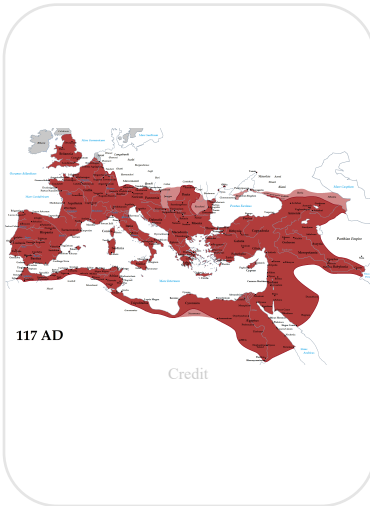
John Huelsenbeck
Department of Integrative Biology
johnh@berkeley.edu

Introduction

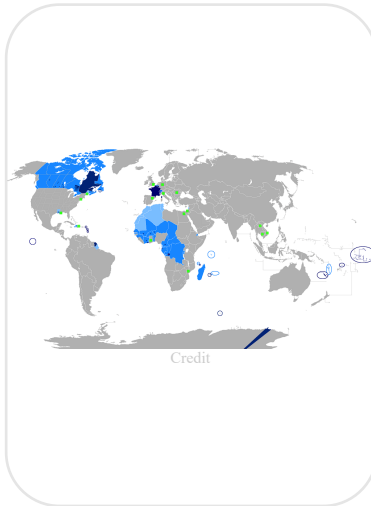
In this paper, we attempt to do the impossible!

Languages

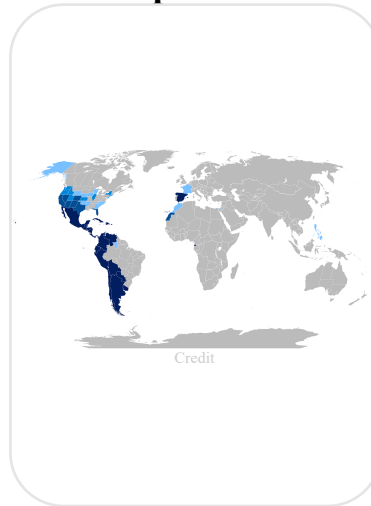
Latin



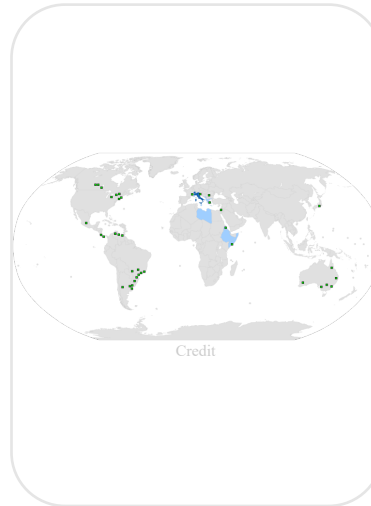
French



Spanish



Italian



Brazilian Portuguese



Portuguese



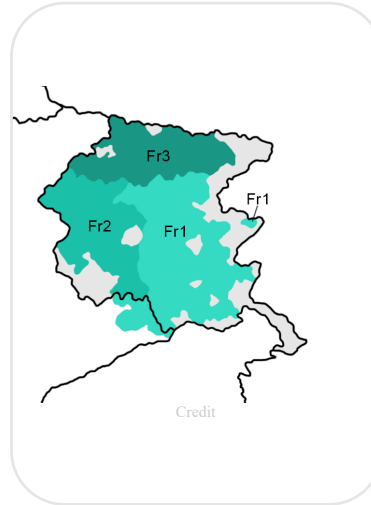
Catalan



Walloon



Friulian

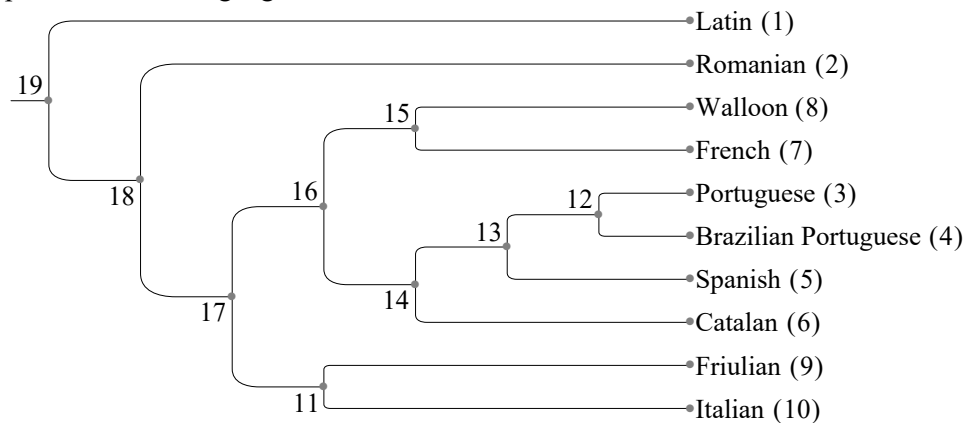


Romanian



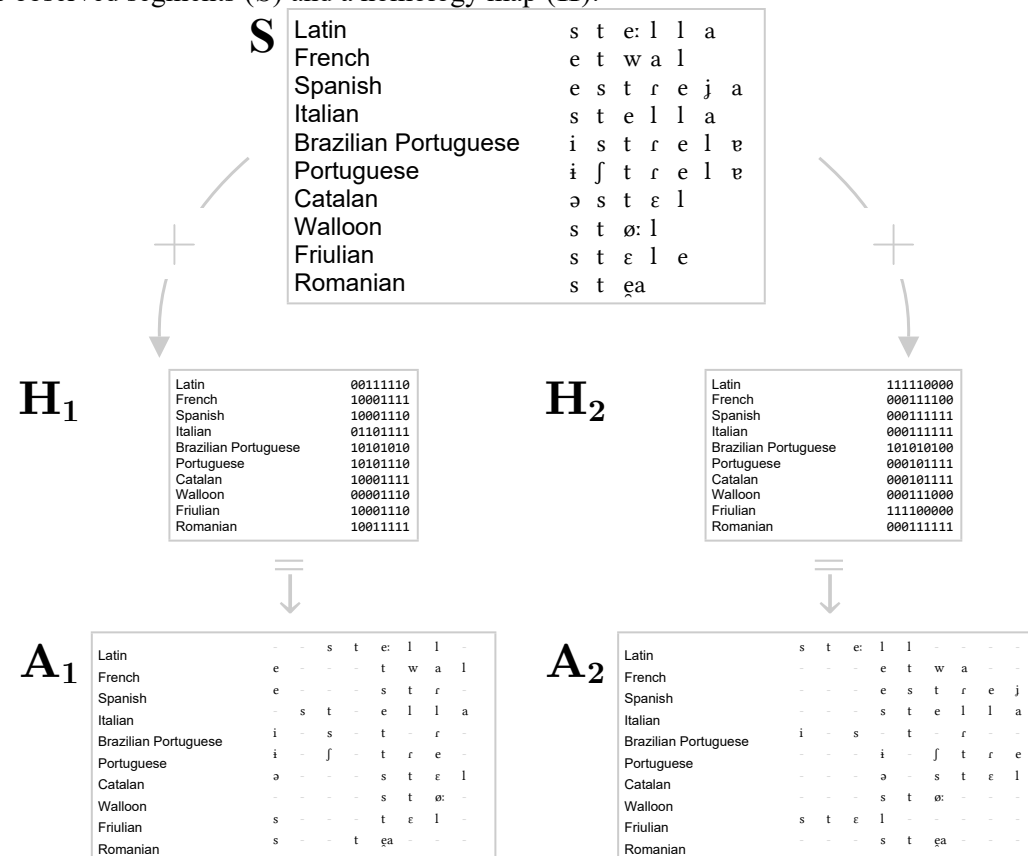
Example Tree

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	o:	4	ʒ	5	ə	6	j	7	o	8	i	9	w	10	ɔ	11	dʒ	12	ɪ	13	j	14	n	15	s	16	u	17	ʃ
18	oj	19	v	20	b	21	t	22	u:	23	y	24	d	25	f	26	e:	27	m	28	a	29	r	30	ɐ	31	œ	32	ɛ				
33	l	34	r	35	x	36	ʎ	37	ʁ	38	i:	39	ɲ	40	ɔ̹a	41	h	42	ẽj	43	õ	44	ã	45	ĩ	46	ẽ	47	tʃ				
48	ɨ	49	ã	50	p	51	z	52	ǫ	53	k	54	ts	55	a:	56	ẽj	57	ʀ	58	ɐ	59	θ	60	ej	61	ɐj	62	õ				
63	ɛ:	64	ɥ	65	ø:	66	ɫ	67	c	68	ɛj	69	ø	70	ɛ̹a	71	k ^w	72	ɣ	73	aj	74	g	75	ẽ	76	ɳ	77	g ^w				
78	ẽɥ	79	ɑ	80	β	81	ʃ	82	ẽ	83	tʃ	84	ɔ:	85	ũ	86	ɭ	87	ç	88	ɑ:	89	ʝ	90	œ̃	91	au̹	92	ʊ̹				
93	ẽw̃																																

Partition Assignments

Basic Rules

1 Nasal Vowel

ẽ ĵ ċ ă ĩ ẽ ă ǵ ǫ ẽ ẹ̃ ̣ ̣ ̣ ̣ ̣ ̣ ̣

2 Vowel

e ɔː ə o i ɔ ɪ u oɪ uː y eː a ɐ æ ɛ iː ʊɑː ɪ aː ɒ eɪ ɛɪ ɜː ɛɪ ø ɛɑː aɪ ɑː ɔː ɑː aʊ ʊ

3 Nasal Consonant

n m j r j

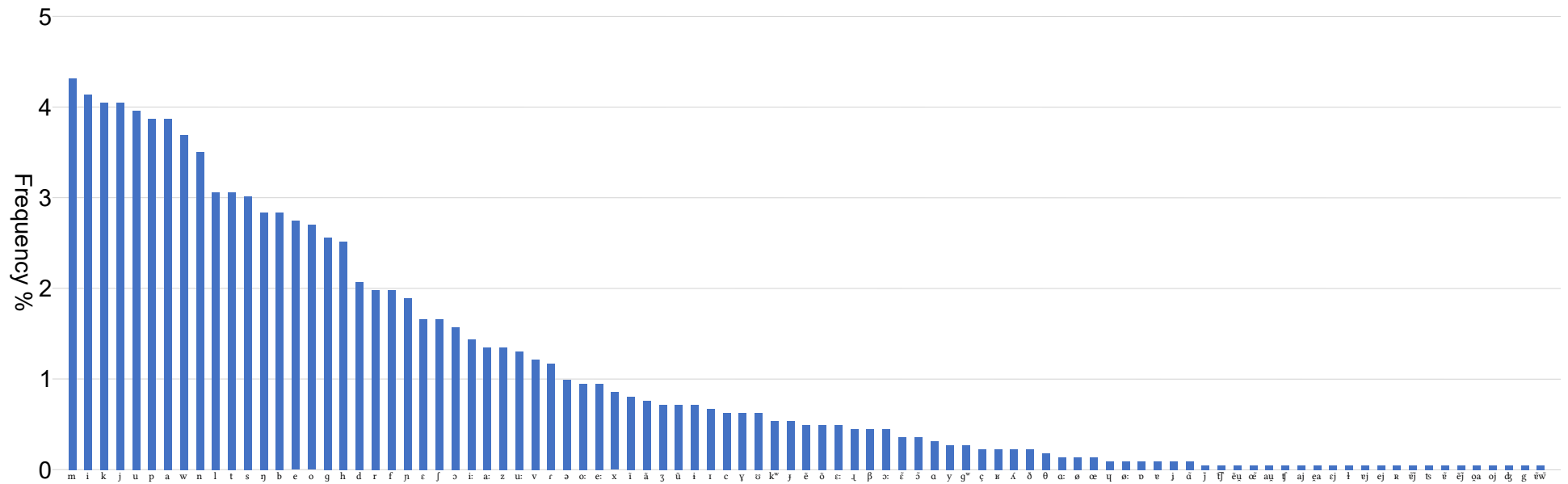
4 Non Sylabic Sonorant

w j l r

5 Consonant

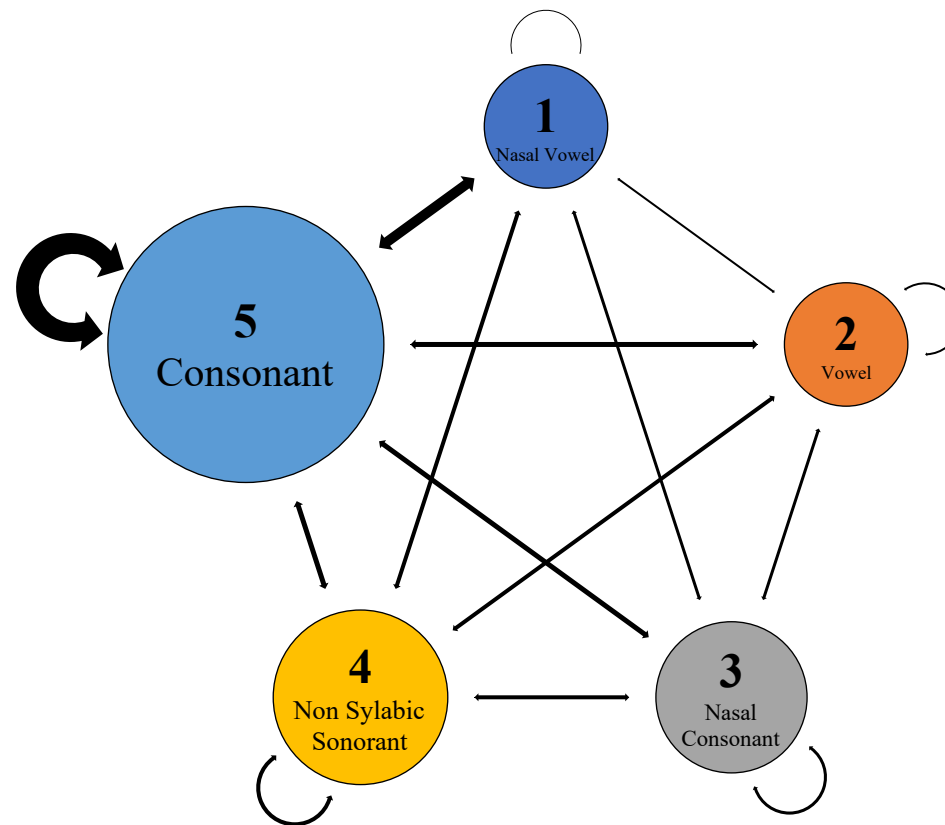
g z j d s f v b t d f r x ʌ ɤ h ʈ p z ʁ k t s r θ ɥ ɫ c k^w ɣ g g^w β ʃ ʈ ɹ ɕ

Prior Segment Frequencies

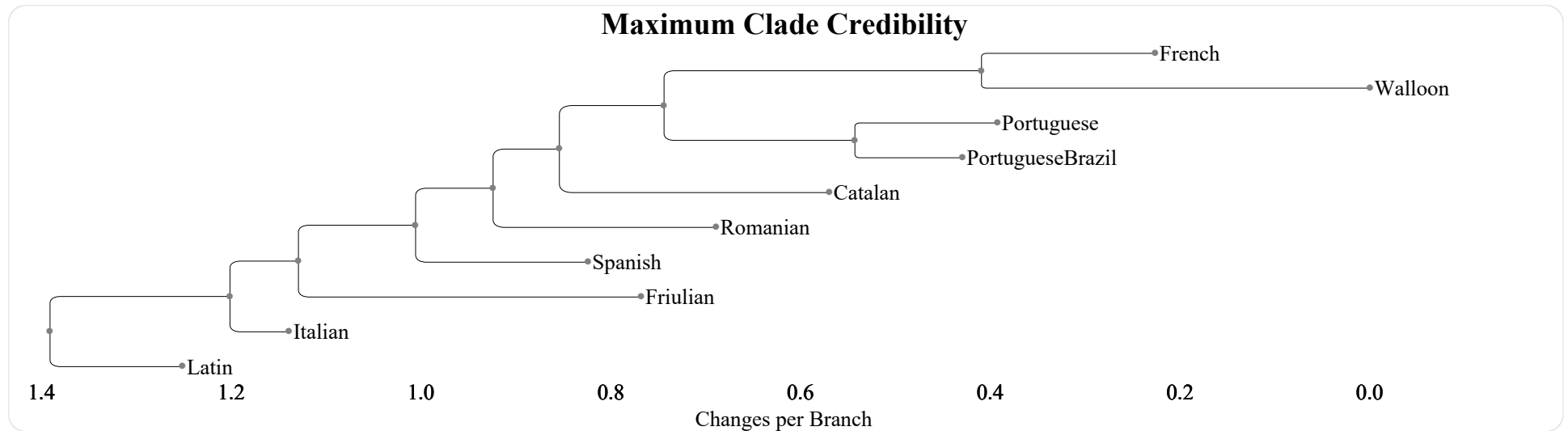


Transition Rates Between Partitions

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



Questions

