

# Geographic Patterns of R1b in the British Isles – Deconstructing Oppenheimer

Kevin D. Campbell

## Abstract

Stephen Oppenheimer's book, *The Origins of the British—A Genetic Detective Story*, references a clan nomenclature which is not explicitly defined in the text nor linked to the underlying data. This paper attempts to understand Oppenheimer's analysis while incorporating results from subsequent clan testing to hypothesize the haplotype definitions for Oppenheimer's R1b sub-clans.

## Introduction

The essence of genetic genealogy is to understand where we've come from. However, many of the recent papers and books have missed opportunities to provide strong links relating genetics to regional locations. Specifically, two recent books by Oxford professors—*Blood of the Isles: Exploring the Genetic Roots of Our Tribal History*, by Bryan Sykes (2006), and *The Origins of the British—A Genetic Detective Story*, by Stephen Oppenheimer (2006) address genetics in the British Isles, but both omit important genetic information. Both books are based upon the analysis of thousands of DNA samples collected in the British Isles, but in each case, critical elements of the analysis are left unpublished—elements that make the analysis difficult to follow and almost impossible for others to confirm the authors' conclusions independently.

In the case of Sykes' work, *Blood of the Isles* was targeted to the general population and written in the manner of a popular work of non-fiction. In contrast, Stephen Oppenheimer's book *The Origins of the British* synthesizes historical, anthropological, archaeological, linguistic, and genetic evidence into a cohesive set of conclusions.

While Oppenheimer chose to include the term "genetics" in the title of his book, it comprises only a part of his overall analysis. However, by leading with genetics, Oppenheimer owes a certain level of traceability to the reader to allow a thorough and detailed review of his analysis.

To his credit, Sykes has provided his samples underlying *Blood of the Isles* for examination, but he failed to pro-

vide a detailed analysis of those samples for the reader. For example, Sykes does not fully describe his "clan" system in adequate detail. A "clan" is a group of individuals with closely matching Y-STR haplotypes, to which a fanciful name has been assigned by the author of the book. However, Sykes does not tell us which haplotypes are included in each clan.

My analysis of Sykes' clans, which included the determination of the probable clan definitions, was published in the last issue of this journal (Campbell, 2007).

In this article I will attempt to achieve the same general result for Oppenheimer's book. In particular, the present article will attempt to provide some insight into the probable definitions of Oppenheimer's clans.

## Methods

Sykes' and Oppenheimer's analyses have both similarities and differences that affect how one might approach reverse-engineering each. Both authors chose to coin "clan names" as shorthand monikers for genetic groups derived from their analysis. While these clan names provide convenient shorthand for mass market books, serious researchers want to see the genetic definitions of these groups.

Sykes' work is based upon original data collection (primarily via blood samples), and he has published his full dataset on his web site for use by other researchers.<sup>1</sup>

Oppenheimer's study is not based upon new genetic data, but rather is based upon a re-analysis of previously published information. Oppenheimer uses five key sources for his British data: The studies of Capelli (2003), Wilson et al (2001), Weale et al (2002), and Hill et al (2000), and data provided by D. Faux and J. Wilson related to the Orkney and Shetland Islands, one

---

Address for correspondence: [Campbell@alum.mit.edu](mailto:Campbell@alum.mit.edu).

Received: July 12, 2007; accepted: August 30, 2007.

---

<sup>1</sup> [http://www.bloodoftheisles.net/OGAP\\_yDNA.pdf](http://www.bloodoftheisles.net/OGAP_yDNA.pdf)

of which (Capelli's) is available on the Internet.<sup>2</sup> Collectively from these sources, Oppenheimer compiled a composite dataset containing 3,084 samples, "though by far the largest body of data in the composite British Isles dataset was collected by Christian Capelli and his colleagues."<sup>3</sup>

Since the vast majority of Oppenheimer's data came from Capelli, the first step in the method was to understand the nature and limitations of Capelli's dataset.

A second part of my method resulted in the identification of the 16 R1b clans that Oppenheimer uses in his analysis. Like Sykes, Oppenheimer does not completely disclose the Y-STR haplotypes that he uses to define these clans. This part of the analysis was done by comparing the reduced six-marker haplotypes and assigned clan designations from Ron Scott's web site (Scott, 2007).

## Results

### *Reviewing Capelli (2003)*

Though Capelli's study, "A Y chromosome census of the British Isles," was published in 2003, only recently has the full underlying dataset been made available.<sup>4</sup> While I will not attempt to summarize the entire analysis here, several points are worth noting. First, the methods of interest to us here can be summarized as follows.

Capelli collected 1,772 Y chromosome samples from 25 predominantly small urban locations in the British Isles. For each sample, Capelli genotyped six Y-chromosome microsatellites (DYS019, 388, 390, 391, 392, and 393) to identify haplotypes. The geographic locations used to sample the populations in Capelli's original study are shown in **Figure 1**.

To understand Capelli's published dataset, the 1,772 British Isles samples were coded by haplogroup using Whit Athey's improved Bayesian haplogroup calculator (Athey, 2006).<sup>5</sup>

The haplogroup mapping results from the calculator and key counts of Capelli's data used in Oppenheimer's study are shown in **Table 1**. The haplogroup results are shown as rows while the column counts are derived from the individual datasets.

It can be seen from this table that the vast majority of Oppenheimer's data is from the Capelli dataset. 71% of his overall data and 85% of that from the British Isles

**Table 1. Summary of Capelli and Oppenheimer Datasets**

	Capelli		Oppenheimer		B.I. Percent Capelli
	British Isles	Non-British Isles	British Isles	Non-British Isles	
E3a	3	-			
E3b	46	12	47		98%
G	21	7			
H	5	1			
I	266	126	336		79%
J	58	14			
L	12	-			
N	25	9			
Q	80	19			
R1a	114	86	126		
R1b	1,142	159	1,511	436	76%
<b>Total</b>	<b>1,772</b>	<b>433</b>	<b>2,082</b>	<b>1,002</b>	<b>85%</b>
<b>Full Data</b>	<b>2,205</b>		<b>3,084</b>		<b>71%</b>

can be traced back directly to Capelli. Oppenheimer has acknowledged this heavy reliance in his book.<sup>6</sup>

Given this reliance, it is clear that Oppenheimer's genetic analysis is based upon the six microsatellites included in the Capelli data. 67 unique R1b haplotypes were extracted from the dataset and are included in **Appendix A**. These 67 haplotypes subsume the 1,301 R1b Capelli samples shown in **Table 1**, and this data represents 76% of all the data used by Oppenheimer in his analysis of R1b migration patterns.

### *Oppenheimer Clans*

Following the same approach as for Sykes' data in Campbell (2007), my next step was to look for patterns that I could link to conclusions listed in the text. However, Oppenheimer has already identified patterns in his analysis. These patterns, or clan groups as he calls them, were identified for haplogroups R1b, I, and R1a. The present article focuses only on R1b. For R1b, Oppenheimer identified 16 clan groups of haplotypes that comprise this haplogroup. He labels these as R1b-2, R1b-3, etc., though some of these groups are further divided into sub clusters (R1b-2a, R1b-2b, etc.).

The missing piece in Oppenheimer's study is the definition of these clusters in terms of the underlying microsatellites. Nowhere in his book are these clusters fully specified.

Though one approach to determining the cluster/clan definitions could be a bottom-up analysis of the data listed in **Appendix A**, essentially reverse-engineering Oppenheimer's work, another strategy was selected. Simply put, since Oppenheimer's genetic clans are

<sup>2</sup> Oppenheimer (2006), Chapter 3, footnote 41.

<sup>3</sup> Ibid, p. 435.

<sup>4</sup> Capelli (2003) dataset located at:

<http://freepages.genealogy.rootsweb.com/~gallgaedhil/Capelli.htm>

<sup>5</sup> Whit Athey's Haplogroup Calculator,

<http://www.hprg.com/hapest5/>

<sup>6</sup> Oppenheimer (2006), p. 123.

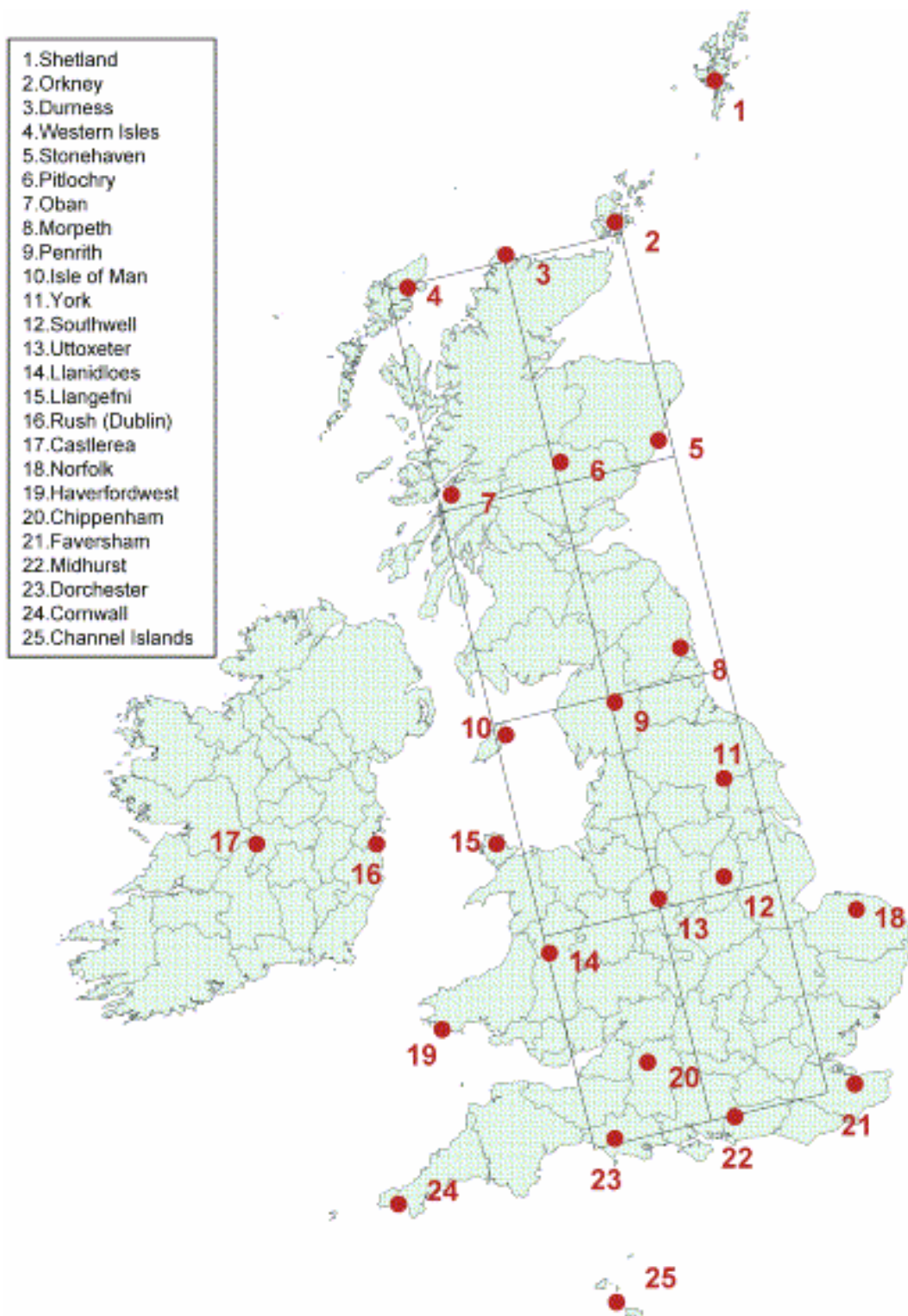


Figure 1. Locations Used for Capelli's Original Data Collection

**Table 2. Oppenheimer Genotypes with Associated Capelli Microsatellites and Intra-Clan Differences Highlighted**

#	Op Subclade	HG	Ysearch	Microsatellites					
				393	390	19*	391	388	392
1	R1b-2b	R1b1c*	3QQDV	12	24	14	11	12	13
2	R1b-2b	R1b1c	YNWV9	12	24	14	11	12	13
3	R1b-2b	R1b1c*	NYT8Z	12	26	14	11	12	13
4	R1b-4	R1b1c	PEZBS	13	25	14	10	12	11
5	R1b-8	R1b1c9*	23UDR	13	23	14	11	12	13
6	R1b-8	R1b1c9*	BGPD5	13	23	14	11	12	13
7	R1b-8	R1b1c9*	7VMB5	13	23	14	11	12	13
8	R1b-8	R1b1c9*	UFBVM	13	23	14	11	12	13
9	R1b-8	R1b1c9*	35UQ4	13	23	14	11	12	13
10	R1b-8	R1b1c	T553A	13	23	14	11	12	13
11	R1b-8	R1b1c*	ZGFUS	13	23	14	10	12	13
12	R1b-8	R1b1c9*	K7T9Y	13	23	14	10	12	13
13	R1b-8a	R1b1c9*	7FZUQ	13	23	14	12	12	13
14	R1b-8a	R1b1c9*	28UTA	13	23	15	12	12	13
15	R1b-9	R1b1c*	MU26C	13	24	14	10	12	13
16	R1b-9	R1b1c9*	6MEY3	13	24	14	10	12	13
17	R1b-9	R1b1c	PJ5DC	13	24	14	10	12	13
18	R1b-9	R1b1c*	GDQ6F	13	24	14	10	12	13
19	R1b-10	R1b1c*	2DSVM	13	24	14	11	12	13
20	R1b-10	R1b1c*	8NQXQ	13	24	14	11	12	13
21	R1b-10	R1b1c*	U7VY3	13	24	14	11	12	13
22	R1b-10	R1b1c	RE7TY	13	24	14	11	12	13
23	R1b-10	R1b1c	BUAJW	13	24	14	11	12	13
24	R1b-10	R1b1c*	PRMKT	13	24	14	11	12	13
25	R1b-10	R1b1c6	QZ8NP	13	24	14	11	12	13
26	R1b-10	R1b1c	MYC6B	13	24	14	11	12	13
27	R1b-10	R1b1c*	DZ7XA	13	24	14	11	12	13
28	R1b-11	R1b1c*	H3EQG	13	25	14	10	12	13
29	R1b-11	R1b1c9*	93UC5	13	25	14	11	12	13
30	R1b-11	R1b1c9*	CN9GE	13	25	14	12	12	13
31	R1b-12	R1b1c*	H7G9X	13	23	15	11	12	13
32	R1b-13	R1b	Y2HKM	13	24	15	10	12	13
33	R1b-13	R1b1c	FYMNA	13	24	15	11	12	13
34	R1b-13	R1b1c9*	R464G	13	24	15	11	12	13
35	R1b-14a	R1b1c9*	CRTJ5	13	23	14	11	12	15
36	R1b-14a	R1b1c*	A2RB3	13	24	14	11	12	14
37	R1b-14a	R1b	3A5MS	13	25	14	11	12	14
38	R1b-14a	R1b1c7	E2M4N	13	25	14	11	12	14
39	R1b-14b	R1b1c9*	DYYG8	13	23	14	11	12	14
40	R1b-14c	R1b1c7	WEYBZ	13	24	14	11	12	14
41	R1b-15a	R1b1c	S286F	14	25	14	11	12	13
42	R1b-15a	R1b1c	S8XD5	14	25	14	11	12	14
43	R1b-15b	R1b1c	2FGYD	14	24	14	11	12	13
44	R1b-15b	R1b1c*	8CZYC	14	24	14	11	12	13
45	R1b-15c	R1b1c9b	ZV3N9	14	23	14	10	12	13
46	R1b-16	R1b1c*	G6RYG	13	24	16	11	13	13
47	R1b-16	R1b1c*	FQ7HW	13	24	14	11	13	13
48	R1b-16	R1b1c*	B4XDU	13	25	14	11	14	13

apparently based primarily on six microsatellites, it was decided to look at how he typed specific participants to attempt to deduce the R1b cluster definitions from their results.

While attempting to collect information on the Oppenheimer clan definitions, another researcher, Ron Scott, decided to compile the same information, from personal communication with participants who had ordered their Oppenheimer clan determinations from a commercial company. Since Ron Scott's compilation was readily available online, this data was used to help identify the Oppenheimer clan definitions<sup>7</sup>

Analysis of the results of Oppenheimer's genotyping has been illuminating. When Oppenheimer Clans are viewed on Ron Scott's Web site as series of 12- or 25-marker haplotypes, there does not appear to be any obvious pattern among the Clan designations. However, when each participant's markers are reduced to the six microsatellites present in the underlying Capelli/Oppenheimer dataset, a definite pattern begins to emerge – i.e. a unique combination of these six markers seem to result in a unique Oppenheimer cluster.

**Table 2** shows the Oppenheimer Clan results from Ron Scott's web site, reduced to the six Capelli microsatellites. In this table, alleles that differ from other clan results are shaded.

It should be noted that because there are fewer Oppenheimer clusters than haplotypes in the Capelli dataset (and fewer than the possible number of combinations of six markers, by necessity), an Oppenheimer cluster must span more than one unique combination of six markers. Or stated another way, since Oppenheimer partitions R1b into only 16 groups, some groups must contain more than one of the 67 haplotypes listed in **Appendix A**.

When looking at Oppenheimer's empiric results in light of Capelli's underlying markers, several conclusions are evident from **Table 2**.

First, no six-marker haplotypes are split among two or more Clans -- i.e., each haplogroup maps into one and only one clan designation. This supports the hypothesis that Clan designations are based primarily on these six markers.<sup>8</sup>

<sup>7</sup><http://freepages.genealogy.rootsweb.com/~ncscotts/Y-DNA/Oppenheimer%20Clan%20Test.htm>

<sup>8</sup> While this statement was true for a long time, a recent empiric posting provides one contradiction in the 48 observations included in Table 2. i.e., Samples #36 and #40 are typed as R1b-14a and R1b-14c, respectively, but contain the same six marker haplotype. The author suspects that this discontinuity is due to lab error but this discrepancy is noted for the reader so they can weigh this anomaly accordingly.

Second, Oppenheimer clan families (e.g., R1b-8 & 8a; R1b-14 a/b/c, R1b-15 a/b/c, etc.) seem to be generally separated by a single step mutation of a single marker. For example, R1b-8 and R1b-8a seem to be differentiated by DYS391 being 11 or 12, while R1b-15a/b/c seem to be differentiated by DYS390 being 25/24/23.

Third, even with the small sample size collected by Ron Scott, the most frequent cluster (R1b-10) and the second most frequent cluster (R1b-8) clan results mirror those found by Oppenheimer in his study.

A summary of Oppenheimer's R1b Clan Tree is reprinted as **Figure 2**. In this figure, the estimated time of branching, the standard deviation, and the number of samples of each clan that were present in his dataset are extracted from various footnotes throughout the Oppenheimer's book. The corresponding six-marker haplotypes are also included where possible.

In Oppenheimer's Analysis, the Atlantic Modal Haplotype (i.e., Ruy or R1b-10) splits from R1b-9 (i.e., Roy) about 9,800 years ago. Oppenheimer found that the second most prevalent group, Haplotype R1b-8 (Clan Rob), branched off later from R1b-10.

When viewed in the context of the aforementioned microsatellites, one can also see how Oppenheimer might draw this conclusion. **Table 3** shows this specific progression of R1b Clans proposed by Oppenheimer in his book.

**Table 3. Haplotype Progression Suggested by Oppenheimer's Analysis**

		Microsatellites					
Clan	Name	393	390	19*	391	388	392
R1b-9	Rox	13	24	14	10	12	13
R1b-10	Ruy	13	24	14	11	12	13
R1b-8	Rob	13	23	14	11	12	13

The haplotype progression shown in **Table 3** further reinforces the conclusion that Oppenheimer's analysis used these microsatellites. The haplotype sequence shown in **Table 3** is logical and follows Oppenheimer's sequence. The haplotype sequence does not support other progressions such as R1b-9 → R1b-8 → R1b-10 or R1b-10 → R1b-9 → R1b-8 that would contradict Oppenheimer's conclusions.

As a final check, the author attempted to recreate several of the Oppenheimer's Clan maps included in his book. In **Figures 3a and 3b**, the data for the



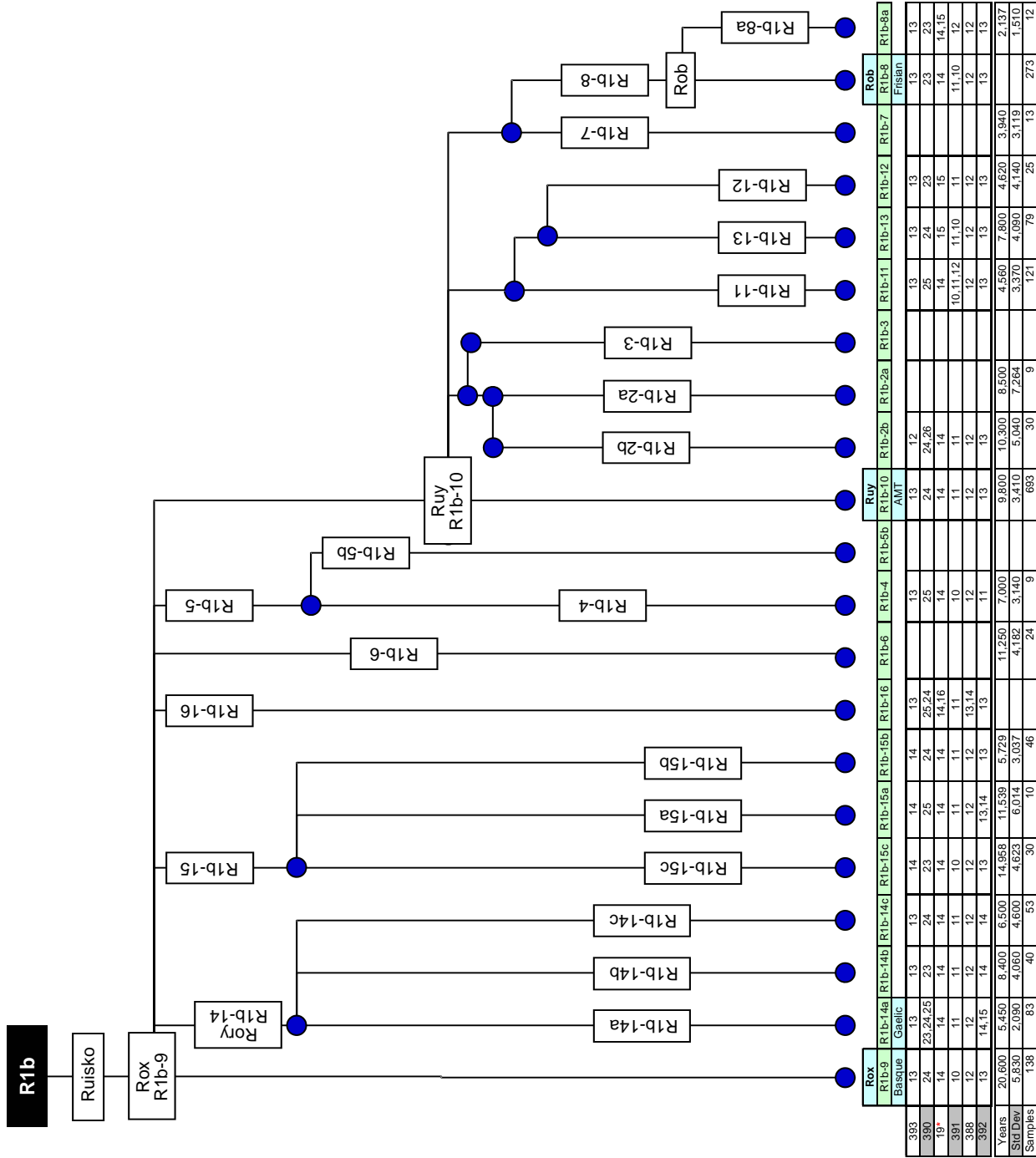


Figure 2. Oppenheimer's Clan Structure with Empiric Y-STR Results

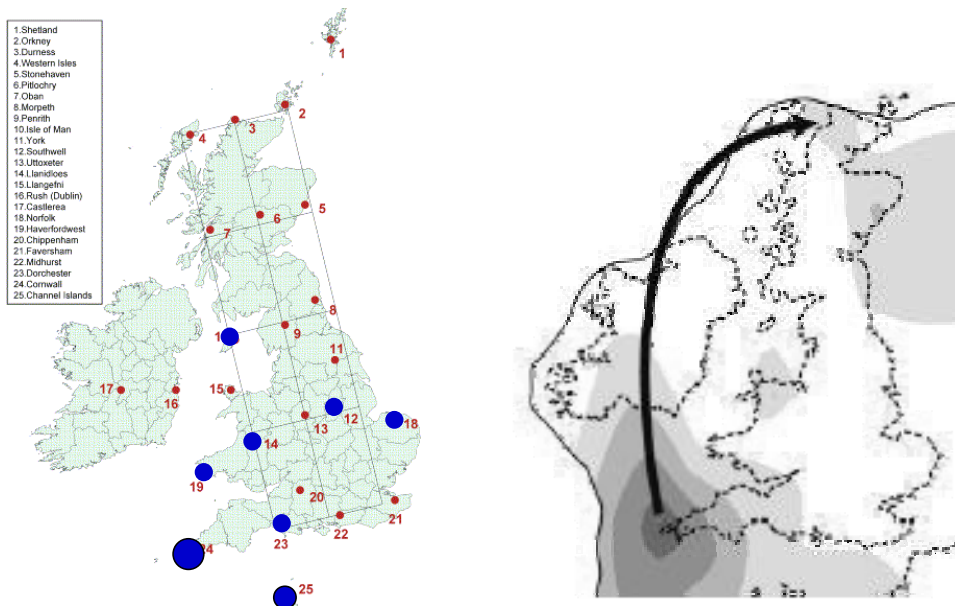


Figure 3a. Comparison of Capelli Samples with Oppenheimer Clan R1b-15c (Size indicates the relative number of observed samples in the Capelli dataset)

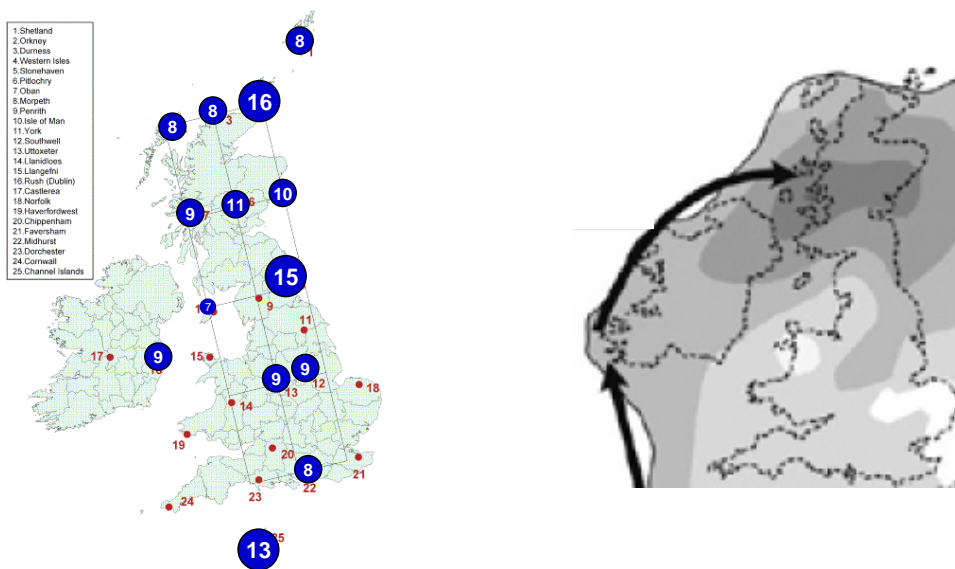


Figure 3b. Comparison of Capelli Samples with Oppenheimer Clan R1b-9 (Numbers indicate number of observed samples in the Capelli dataset)

hypothesized clans R1b-15c and R1b-9 were plotted on Capelli's map of the British Isles with circles sizes representing the number of samples observed by Capelli. Data for these figures is from the underlying Capelli data shown in **Appendix A**.

There are some limitations to this analysis. For example, (1) the known *ex post facto* clan samples shown in **Table 2** are very small, (2) not all of Capelli's known haplotypes have been genotyped into Oppenheimer Clans, and (3) while significant, only 85% of Oppenheimer's R1b British Isles data is attributable to Capelli's underlying dataset in the first place.

These caveats notwithstanding, the author believes that **Figure 3** further reinforces the assertion that the insights into Oppenheimer's clan nomenclature can be deduced when *ex post facto* results are compared to Capelli's dataset.

## Conclusions

The analysis presented in this paper tends to confirm the hypothesis that Oppenheimer's Clan system is based upon the six microsatellites presented in the data of Cristian Capelli. Though only a small number of samples have been genotyped according to this system since the book has been published, these samples offer ample evidence to speculate on the haplotype signatures of specific Oppenheimer Clans. This paper has speculated on Haplotype matches for 16 of 21 Clans and Sub-Clans depicted in Figure 2. From the 15 for which we have both proposed haplotypes definitions and statements by Oppenheimer of the frequency of these Clans in his full dataset, we see that this analysis suggests that the sub-clans listed in **Table 2** account for 84% of all the R1b data used by Oppenheimer.<sup>9</sup>

---

## Electronic Data Base Information

Capelli C, et al.(2003) dataset:

<http://freepages.genealogy.rootsweb.com/~gallgaedhil/Capelli.htm>

Oxford Genetic Atlas Project (OGAP), data from Sykes (2006):

<http://www.bloodoftheisles.net/results.html>

John McEwan's R1b Haplotypes:

<http://www.geocities.com/mcewanjc/p3modal.htm>

Ron Scott's Database of Oppenheimer Clan Results:

<http://freepages.genealogy.rootsweb.com/~ncscotts/Y-DNA/Oppenheimer%20Clan%20Test.htm>

Whit Athey's Haplogroup Predictor

<http://www.hprg.com/hapest5/>

---

## References

Athey TW (2006) Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach, *J Genetic Genealogy*, 2:34-39.

Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MP, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N, Goldstein DB (2003) A Y chromosome census of the British Isles. *Curr Biol*, 13:979-984.

Oppenheimer, Stephen. *The Origins of the British—A Genetic Detective Story*. Constable and Robinson, New York (ISBN 1-84529-158-1).

Sykes B (2006) *Blood of the Isles: Exploring The Genetic Roots of Our Tribal History*. Bantam, London (ISBN-10:0593056523)..

---

<sup>9</sup> 1,642 samples are accounted for out of the full 1,947 R1b sample data set. i.e., 1,511 plus 436 (See Table 1)



