David N. Ewing, M.D.
June 6, 2008

Modal Haplotypes

If you already understand the concept and implications of modal haplotypes, skip the first several paragraphs and have a look at the table below, which compares the various Ewing Group modals and some other modal haplotypes of interest. If not, then get yourself up to speed by reading the next few paragraphs.

A "**haplotype**" is just a list of the numbers of repeats at a series of markers. In the Ewing project we are mainly considering 37 markers, and we have "37-marker haplotypes" for most of our participants. This means that we can report the results of each participant by giving a list of 37 numbers. The list for each participant is his haplotype. We can also speak of a subset of the marker results on a participant as his haplotype, but in that case to avoid confusion we must be careful to specify what subset we are speaking about. For example, in *Y-DNA Article 12* I spoke of "Capelli 6-marker haplotypes," which means that I was talking about just the six markers that Capelli had used in his study. In the Ewing project, when we say "haplotype" without any adjectives, we mean "FtDNA 37-marker haplotype."

"**Modal**" is a concept from mathematics that is so simple that some folks find it hard to understand. The modal value of a list of numbers is just the number that appears most frequently in the list. For example, the modal value of the list 1, 1, 3, 4, 16 is 1. As you can see, the modal is not the mean or average (which would be 5) or the median (which would be 3). No calculations are necessary--to find the modal value of a list you just count how many times each number appears on the list and the number that appears the greatest number of times is the modal. Not every list of numbers has a modal value. Consider the list 1, 2, 3, 4. Which number appears most frequently? Why, none of them--each appears only once, so this list has no modal. How about 10, 10, 11, 11, 10? Yep, the modal is ten. And how about 10, 11, 10, 11? Welp, there are two tens and two elevens, so there is no modal.

So what is a "**modal haplotype**?" Remember that a haplotype is just a list of numbers, so a modal haplotype is just a list of numbers. But rather than it being a list of numbers of repeats at a series of markers in an individual man, it is a list of modal values for the haplotypes of a group of men at a series of markers. Huh? Let's make up some 3-marker haplotypes for a group of three men and construct a 3-marker modal haplotype for them. Let the men have these 3-marker haplotypes:

Larry	10, 11, 17
Moe	10, 12, 18
Curly	10, 11, 18

The modal value for the first marker is easy; everyone has ten, so ten is the most frequently appearing value and therefore the modal is ten. The modal values for the next two markers are no more difficult to find; they are 11 and 18, respectively. The modal haplotype for this group of three men is 10, 11, 18. As you can see, Curly's haplotype exactly matches the group's modal haplotype and

each of the other men have a single difference from the modal at different markers.

Implications of the Modal Haplotype

So, who cares? What does a modal haplotype have to do with anything? Well, the fact is that the modal haplotype of a group of related men is very likely to have been the haplotype of their common ancestor. I hope you are asking why this should be so. Consider that mutations are rare events. If we are studying 37 markers that have an average mutation rate of 0.4% (one in 250 generations), it doesn't become more likely than not that we will see a mutation somewhere in the 37 markers until the fifth great grandson. And if we test ten fifth great grandsons of a single man, who have no more recent common male ancestor, the odds are about 70% that we will find a total of somewhere between seven and ten mutations among them. Since mutations occur at random, these mutations will be distributed randomly among the men we test. Some of them won't have inherited any mutations, maybe some will have inherited a couple, but on average they will have inherited one apiece, and we will expect all of the men to have their mutations a different markers for the most part. So one man will have a mutation over here, another will have a mutation over there, etc. Can you see what this will make the table of haplotypes look like? For each marker, practically everybody will have the same value and only one or two will have a different (mutated) value. So what is the modal in this case? As always, the modal is the most frequently appearing value, that is, the un-mutated value. But this is just the value that the fifth great granddad started with, no? This is why we are so interested in modal haplotypes: they are our best estimate of what the haplotype must have been of an ancestor we cannot test. What is more, we don't even have to know who he was or when he lived. If we have enough of his offspring and are confident of the mutation rates, we can calculate how many generations ago he is likely to have lived.

Comparison of Selected Modal Haplotypes

All of the modal haplotypes below are compared with the "Ewing modal haplotype." The Ewing modal haplotype was constructed as described above by using all of the Ewing haplotypes in the project. As it happens, there is a large enough number of related men in the Ewing project that the modal haplotype of the subset of related men is also the modal of the entire group. The fact that we have used the shorthand label "Ewing" in the tables to identify this modal haplotype is in no way intended to imply that Ewing men who do not match it closely are somehow not genuine Ewings. Indeed, constructing a modal haplotype for all of the project participants together obscures the fact that some of the smaller subsets of men who are related to one another but not to the large group of related men have modal haplotypes pointing to common ancestors of their own. This is true of Groups 6 and 9, and of a couple of subsets of Group 8, and I suppose it would be true of Group 2, except that it doesn't make any sense to speak of a modal haplotype when we have 37-marker data on only one man in that group. On the other hand, Groups 1, 3, 4 and 7 all have the same modal

haplotypes, and none of their members are very far from this modal. This demonstrates that they have a common ancestor, even though we haven't been able to establish who he may have been. Except for a couple of small clusters within it, Group 8 is a heterogeneous group of men who are unrelated to one another in a genealogical time frame. They are all in Haplogroup R1b1, which means they do have a common male ancestor, but he most likely lived something like 12,000 years ago during the last glacial maximum. Have a look at the table, and then we can say more about it.

Selected Modals Compared to the Ewing Modal Haplotype

ID	DYS393	DYS390	DYS19/394	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389-1	DYS392	DYS389-2	DYS458	DYS459a	DYS459b	DYS455	DYS454	DYS447	DYS437	DYS448	DYS449	DYS464a	DYS464b	DYS464c	DYS464d	DYS460	GATAH4	YCAIIa	YCAIIb	DYS456	DYS607	DYS576	DYS570	CDYa	CDYb	DYS442	DYS438
Ewing	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 1	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 3	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 4	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 7	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 5	13	25	15	10	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
Group 5-1	13	25	15	10	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	35	37	11	12
R1b1b2e	13	25	14	11	11	13	12	12	12	13	14	29	17	9	10	11	11	25	15	18	30	15	16	16	17	11	11	19	23	17	16	18	17	38	39	12	12
R1b	13	24	14	11	11	14	12	12	12	13	13	29	17	9	10	11	11	25	15	19	29	15	15	17	17	11	11	19	23	15	15	18	17	37	38	12	12
Group 8	13	24	14	10	11	14	12	12	12	13	13	29	18	9	10	11	11	25	15	19	29	15	15	17	17	11	11	19	23	16	15	18	17	36	38	12	12
R1bSTR47	13	24	14	10	11	14	12	12	12	13	13	30	18	9	10	11	11	25	15	19	30	15	15	17	17	11	12	19	24	16	15	18	17	37	38	12	12
Group 6	13	24	15	11	11	15	12	12	12	12	13	28	18	9	10	11	11	25	15	19	29	15	15	17	17	11	12	19	23	16	14	18	17	36	38	12	12
Group 9	15	22	16	10	14	14	11	13	11	12	12	29	18	8	10	11	11	22	15	20	31	11	15	15	15	11	11	19	21	14	15	17	18	33	38	12	10

Distance from reference:	Zero	One	Two	Three+

Table prepared using *Dean McGee's Y-DNA Utility*.

You can see that the overall Ewing modal is identical to the modals for Groups 1, 3, 4 and 7. Group 5 is defined on the basis of differing from the Ewing modal at DYS 391 = 10, and the overall Group 5 modal matches the Ewing modal except at that marker. You might be tempted to make something of the fact that the STR47 and Group 9 modals also have DYS 391 = 10, but they are otherwise so different from Group 5 that I have concluded these are so-called parallel mutations; that is, they are coincidental and not due to inheritance from a common ancestor. Group 5, Part 1 also differs from the Ewing modal and the rest of Group 5 at CDYa/b, and we think this pattern may distinguish the descendants of William?, the progenitor of this sub-group. The modal labeled R1b is actually the Atlantic Modal Haplotype, which is the most common R1b haplotype among western Europeans and is especially prominent along the Atlantic facade. I have

included the R1b1b2e modal[1] and McEwan's R1bSTR47-Scots modals for interest.[2]

I constructed a Group 8 modal haplotype, but this is a little silly. You can make a modal from the haplotypes of any random list of men, but it won't mean anything unless they actually do have an interesting degree of relationship. Since we don't think the men in Group 8 are any more closely related than they are by virtue of the fact that they are all in haplogroup R1b, we would expect the Group 8 modal to match the R1b modal, and I think it probably would if we had a large enough number of men in Group 8. As it is, the Group 8 modal differs from the R1b modal only at four places. Isn't it interesting that one of them is at DYS 391 and that there it and the Group 5 modal match the R1bSTR47-Scots modal? Interesting, perhaps, but I haven't been able to make anything of it. I did not construct a modal for Group 2 because we have 37-marker data on only one man in that Group. If you want to see how his haplotype compares with the Group 9 and Ewing modals, have a look at the Group 2 Results Table (he is genetic distance thirty-nine from the Group 9 modal).

Notice the relative genetic distances between these modals to get some idea about the relative duration since these groups split off from one another. Several of the Ewing groups are so close as to be genetically indistinguishable and we think they may have diverged within a genealogical timeframe, maybe four hundred years ago, or so. They are genetic distance seven from the UiN modal and may share a common ancestor with the rest of R1b1b2e maybe a couple of thousand years ago. They are genetic distance fifteen from the R1b modal and nineteen from the STR47 modal, and may share a common ancestor with them from around 10,000 years ago. The Ewing modal is genetic distance forty-nine from the Group 9 modal (and fifty-five from JD, the only man with 37-marker data in Group 2, shown in the Group 2 Results Table). Groups 9 and 2 are both in Haplogroup I, but are only very distantly related to one another and their common ancestor probably lived at a time near the origin of Haplogroup I, maybe 20,000 years ago. Confidence intervals on these estimates are enormous, so don't put too much stock in the numbers, but you can see the impact of the relative differences. Notice that a genetic distance of 8-10 from the Ewing modal is still in R1b1c7[3] and genetic distance 20 or so is still in R1b, but a genetic distance of 50 is in a completely different haplogroup. For more discussion about this, have a look at *Haplogroups, Haplotypes and Clusters for the Flustered*.

---

[1]The name of R1b1c7 has recently been changed to R1b1b2e. By either name, this is the same as the Ui Niall modal haplotype.

[2] McEwan's other R1bSTR modals and other haplogroup and cluster modals are posted on his *website*, but heads up: he uses a different convention for reporting DYS 389-2. To make the value in his tables compatible with the FTDNA convention, add the numbers he shows for DYS 389-1 and DYS 389-2 and use the sum for DYS 389-2.

[3] Those interested in knowing more about this should read *Y-DNA Article 7*, and John McEwan's article, *What can Y-DNA tell us about the Ewings?* You can see a phylogenetic tree that shows the genetic relationship between these groups graphically at *www.geocities.com/mcewanjc/ewing.pdf*. (You will have to blow this up at least 200-300% to read the individual names.)