npg

ARTICLE

The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis

Brian McEvoy[1], Claire Brady[1], Laoise T Moore[1] and Daniel G Bradley*[,1]

[1]Smurfit Institute of Genetics, Trinity College, Dublin, Ireland

The Vikings (or Norse) played a prominent role in Irish history but, despite this, their genetic legacy in Ireland, which may provide insights into the nature and scale of their immigration, is largely unexplored. Irish surnames, some of which are thought to have Norse roots, are paternally inherited in a similar manner to Y-chromosomes. The correspondence of Scandinavian patrilineal ancestry in a cohort of Irish men bearing surnames of putative Norse origin was examined using both slow mutating unique event polymorphisms and relatively rapidly changing short tandem repeat Y-chromosome markers. Irish and Scandinavian admixture proportions were explored for both systems using six different admixture estimators, allowing a parallel investigation of the impact of method and marker type in Y-chromosome admixture analysis. Admixture proportion estimates in the putative Norse surname group were highly consistent and detected little trace of Scandinavian ancestry. In addition, there is scant evidence of Scandinavian Y-chromosome introgression in a general Irish population sample. Although conclusions are largely dependent on the accurate identification of Norse surnames, the findings are consistent with a relatively small number of Norse settlers (and descendents) migrating to Ireland during the Viking period (ca. AD 800–1200) suggesting that Norse colonial settlements might have been largely composed of indigenous Irish. This observation adds to previous genetic studies that point to a flexible Viking settlement approach across North Atlantic Europe.
European Journal of Human Genetics (2006) **14,** 1288–1294. doi:10.1038/sj.ejhg.5201709; published online 6 September 2006

Introduction

The Vikings or Norse have played a prominent role in Irish history. After first recorded contact in AD 795 they are credited with establishing Ireland's first true urban settlements including modern Dublin, Limerick and Cork.[1,2] It is generally agreed that present day Norway was the source of Ireland's Viking raiders and settlers but despite their apparent historical prominence, other lines of evidence, including the scarcity of Norse place names in Ireland and the paucity of word borrowing from Norse into Irish, argue against any large-scale Viking immigration.[3] In addition, there also appears to have been extensive integration between the Irish and Norse leading to a fusion 'Hiberno–Norse' material culture in areas of Viking settlement. Nonetheless, the inhabitants of Hiberno–Norse towns were distinctively referred to as *Ostmen* (old Norse for 'men of the east') to distinguish them from the rest of Ireland.[4]

The genetic legacy of the Viking age is well characterised for many areas of North Atlantic Europe.[5–8] However, there has been relatively little investigation of their impact on the Irish gene pool. Principal component analysis (PCA) of Y-chromosome allele (haplogroup) frequencies suggested little or no Norse introgression in Ireland, whereas maximum likelihood admixture calculations indicated a substantial 34.8% Norwegian ancestry in a Dublin hinterland sample.[9] However, the confidence

intervals on this estimate encompassed virtually zero to total Norwegian ancestry.

Unlike many areas of the British Isles, Norse kingdoms in Ireland were limited in contiguous geographic extent[1] and centred on now highly urbanised cities making it difficult to reliably identify modern potential Norse-Irish geographic populations. As an alternative approach, we used the patrilineal coinheritance of surnames and Y-chromosomes.[10] Surnames were first adopted in Ireland from ca. AD 900 to 1200, roughly coinciding with the period of Viking activity. During this time, numerous Norse personal names (and nicknames) were introduced onto the island, which later passed into hereditary surnames often by the prefixing of *Mac* or *Ó,* meaning 'son of' and 'grandson/descendent of', respectively. Not every putative Norse surname must necessarily have been founded by a Norse male. However, as a group, current bearers of these surnames have a *prima facie* link to Viking society and may be at least enriched for such ancestry. A previous Irish Y-chromosome[11] study using this approach successfully identified a distinctive paternal population history for Irish groups with surnames of exogenous (English and Scottish) origin. A group of Norman and Norse surnames showed less evidence for an unusual, Y-chromosome history though the sample size was small.

We have assembled a cohort 47 Irish men bearing surnames of putative Norse origin and characterised the Y-chromosome diversity of each through unique event polymorphism (UEP) and short tandem repeat (STR) loci. The relative proportions of Irish and Scandinavian ancestry are examined independently for UEP and STR diversity, using six different admixture estimators to examine the scale and nature of Norse settlement in Ireland and allow a parallel investigation of the impact of method and marker choice on Y-chromosome admixture estimates.

Materials and methods
Samples and genotyping
Norse surnames were identified according to McLysaght,[12] the major general reference for individual Irish surname origins. The number of surnames with a putative Norse origin is relatively low and in some cases there are multiple possible origins for a single surname. For these reasons, the sample population of 47 men came from 26 names thus including multiple representatives of some names as follows (number): Arthur (2), Beirne (2), Bligh (1), Boland (3), Caskey (1), Coll (1), Coppinger (2), Doyle (4), Gohery (2), Hanrick (2), Harold (1), Hendrick (1), Higgins (2), Kells (2), Kettle (1), Loughlin (1), McGetrick (3), McLoughlin (3), Nelson (1), Norris (1), O'Higgins (1), Sugrue (2), Sweetnam (1), Thunder (1), Toner (3) and Tormey (3). Volunteers were recruited through personal contacts or postal request to relevant surname bearers selected from telephone direc-

tories. Most of these came from geographic areas associated with Norse activity in Ireland.

DNA was extracted using a standard phenol/chloroform protocol and each sample genotyped for 13 STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS385A, DYS385B, DYS389I, DYS389II, DYS460, DYS461, and DYS462) in two multiplexes (MS1 and EBF) essentially as previously described.[13] Up to six UEPs (M269, M170, M26, SRY-1532.2 (SRY10831.2), M35, YAP), which divide samples into broader haplogroups,[14,15] were also examined in a hierarchical manner using polymerase chain reaction/restriction fragment length polymorphism assays. Although STR profiles were generated for 47 samples, complete UEP typing was only possible for 43 of these. The haplogroups of the remaining four samples were inferred from their STR profiles by comparison to a database of ~2000 Irish Y-chromosomes[16,17] typed for these STRs and UEPs.

The current data were analysed primarily in concert with published Y-chromosome data for Ireland, Scandinavia and Britain.[9] Norse surname group (NSG) Y-chromosomes were divided into haplogroups as outlined by Capelli *et al*.[9] The single nucleotide polymorphisms M173 and M17 used in the former study were not typed in the NSG. However, M269 and SRY1532.2, respectively, were taken as phylogenetic equivalents for the purposes of comparable haplogroup definition. Virtually all North-West Europe haplogroup R Y-chromosomes appear to either have the derived allele for SRY1532.2 and M17 or M173 and M269 meaning that this treatment introduces minimal error (the full Y-chromosome phylogenetic tree can be seen in Jobling and Tyler-Smith[15]). STR-based analysis was restricted to six marker haplotypes (DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393) to allow maximum comparability between datasets.

Admixture analysis
Under the hypothesis that the NSG is an admixed Irish-Norse population, we attempted to estimate the admixture proportion (m) derived from each parental population. These estimates are generally based on allele frequencies in modern parental and hybrid populations. However, these can fluctuate after the admixture event by genetic drift, migration, mutation or selection. Sample population allele frequencies may also suffer from stochastic sampling error. Several different admixture estimation methodologies are commonly used, many of which try to incorporate some of these potential complications.[18] However, no current admixture method can fully account for all of them. Six of the most commonly employed estimators were used to provide a broad-based assessment of NSG admixture proportions.

Estimates were carried out using UEP and STR diversity independently because of difficulties in combining both for some methods (see Helgason *et al*[6]). This dual approach

allows an investigation of the efficacy of methods based on a small number of discrete alleles (haplogroups) *versus* those from a higher number of alleles (multi-STR loci haplotypes), which have a greater 'noise' risk from recurrent and back mutations. UEP alleles are simply represented by the 11 haplogroups defined in Capelli et al[9] whereas Y-chromosome STR alleles are compound haplotypes of six STR loci. The first parental population (P[1]) was an Irish sample ($n = 119$), whereas a Norwegian sample ($n = 201$) represented the second parental population (P[2]).[9] This data set was chosen to allow strict comparability in sample populations between UEP and STR-based admixtures estimates. A brief description of each admixture estimator examined follows:

(1) m_R[19]: A least-squares estimator that directly uses allele frequencies in calculations. m_R was estimated using ADMIX 2.0.[20,21] Confidence intervals are given by the standard deviation over 10 000 random bootstrap samples.

(2) m_C:[22,23] A later variation on m_R, m_C attempts to incorporate the effects of drift in the hybrid population and stochastic sampling error in the hybrid and parental populations. m_C was calculated using LEAD-MIX[24] ('Likelihood Estimation of Admixture'), Version 2.0 (http://zoo.cam.ac.uk/ioz/software.html).

(3) m_Y[20]: incorporates the effect of mutation since the admixture event. The molecular relationship between alleles was defined as the number of pairwise differences for UEP haplogroups or the sum of the squared difference in allele sizes for STR haplotypes. The date of the admixture event was set at 1000 years ago applying a mutation rate of 2.2×10^{-7} per year for UEPs[25] and 1.38×10^{-5} per year for six-locus STR haplotypes.[26] (The use of alternative rates was not found to significantly affect the results in this case.) Computations were carried out using ADMIX 2.0.[20,21] Confidence intervals were obtained as the standard deviation of 10 000 random bootstrap samples.

(4) m_L[27]: A coalescent-based maximum likelihood method. Ancestral allele frequencies compatible with the observed data are reconstructed and the probabilities of obtaining the current frequencies in the hybrid group are calculated for different parental population contributions. m_L was calculated using LEA[27] ('Likelihood-based Estimation of Admixture') (http://www.rubic.rdg.ac.uk/~mab/software.html) with 55 000 Markov chain Monte Carlo (MCMC) iterations, the first 5000 discarded as burn-in. The median value was taken as the estimate of m_L whereas the 97.5 and 2.5 percentiles provided the 95% credible interval.

(5) m_W:[24] A maximum likelihood approach that emphasises the effect of genetic drift, both in terms of parental population differentiation prior to the admixture event and subsequently in the admixed

and parental populations. It also incorporates the effect of sampling error for each population. m_W estimates were calculated using LEADMIX[24] and are returned bounded by 95% credible intervals.

(6) m_ρ:[6] m_ρ allows for the incorporation of 'private lineages' (alleles not observed/sampled in either parental population possibly owing to mutation since the admixture event) by identifying the closest match (or matches) as an ancestral haplotype(s). Distances between alleles, for the identification of potential ancestor(s) of private lineages, were defined as the sum of repeat unit difference between haplotypes. Computations were implemented using software[6] provided by Agnar Helgason (deCODE Genetics, Reykjavík).

Genetic distances

The relationship of the NSG to 10 geographically defined population groupings in the North Atlantic region was also examined using a genetic distance approach. Norwegian ($n = 201$) and Danish/North German ($n = 190$) population samples were taken directly from Capelli et al[9] while a number of British Isles subpopulations were generated from the same data by combining smaller groupings as follows (n): Ireland (119); Scottish Isles (272); Scotland (178); Wales (196); England North (231); England Midlands (205); England Southwest (125); England Southeast (256).

Once again analysis was conducted independently for UEP and STR markers using the ARLEQUIN package[28] (Version 2.000). Population genetic distances, as linearised Φ_{ST} values, were calculated from the mean pairwise difference between UEP haplotypes. Φ_{ST} significance was ascertained by randomly permuting individuals across populations over 10 000 replicate analyses. The interpopulation Φ_{ST} matrix was summarised in two-dimensional space using the ALSCAL program in the SPSS package (Version 11.0, SPSS Inc.). A similar analysis was conducted for six-STR loci haplotypes using linearised R_{ST} values.
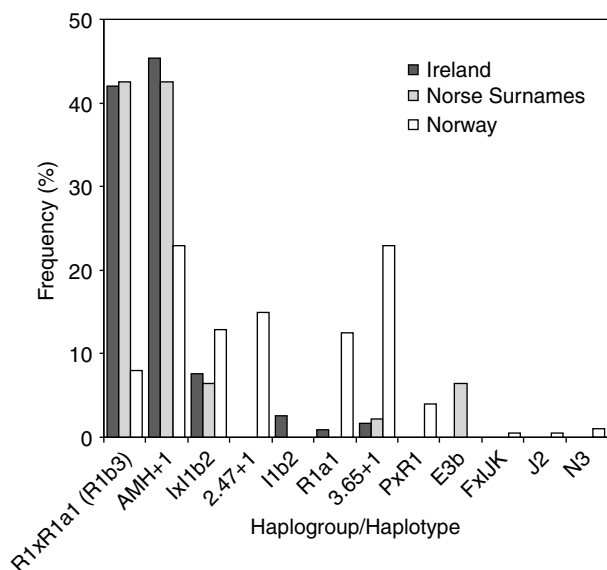
Results

There is a generally strong correlation of Irish surnames with Y-chromosomes indicating shared recent patrilineal ancestry.[17] However, this varies considerably between names, reflecting different origins (single *versus* multiple founders) and the impact of post-foundation male introgression. Thirty-four (13 STR) haplotypes were observed in our cohort of 47 men bearing one of 26 putative Norse surnames. Most haplotypes are either unique or found across different surnames (Genotype data can be found in Supplementary Table 1). However, there is some Y-chromosome haplotype sharing within surnames; for example, in the Arthur, Hanrick and Gohery surnames

both samples from each were identical suggesting a single origin or introduction to Ireland. Conversely, all four Doyle Y-chromosomes are different to each other tentatively pointing to multiple foundations.

Previous studies have identified Y-chromosome lineages that are thought to be indicative of Viking contact and introgression. Two such UEP/STR-defined haplotypes and their one-step STR neighbours ('+1' those haplotypes differing by one STR repeat over all loci) are found at high frequency in Norway (representing the Viking source population) and other areas affected by Viking migrations.[29] These types, termed $2.47+1$ and $3.65+1$, are either absent or at low frequency in the NSG (Figure 1). In contrast, the Atlantic modal haplotype (AMH + 1) within haplogroup $R1 \times R1a$ (R1b3), typical of Ireland and Western Britain, is very common in the NSG. The N3 and $P \times R1$ haplogroups also found in Norway at appreciable frequencies are absent in the British Isles (other than the Orkney Islands)[9] and the NSG. Overall, of the 20 STR haplotypes found in the NSG (based on a truncated 6 STR haplotype for inter-data set comparison) only one is more common in Norway than Ireland, 12 display the converse pattern, whereas seven are unique to the NSG (Supplementary Table 2).

We also examined the occurrence of these putative Viking marker lineages in a general Irish population sample of 796 Y-chromosomes.[16] However, these are not frequent (<0.5%) in the Irish population as a whole nor do they comprise a large fraction of Y-chromosomes in any of

nine smaller geographic Irish subgroupings, indicating no strong regional-specific introgression.

The relationship of the NSG group to 10 North Atlantic European geographic populations was further examined, for UEP and STR diversity separately, using genetic distances (Φ_{ST} and R_{ST} respectively). Multidimensional scaling (MDS) summaries of these pairwise genetic distance markers are shown in Figure 2. Each of the two dimensions are strongly correlated between the different marker systems ($r = \sim 0.9$) and demonstrate a significant differentiation of Norwegian and Irish Y-chromosomes and a markedly closer affinity of the NSG to Ireland. This is particularly evident within relationships based on multi-STR haplotypes, which seem to have a generally greater discriminatory power. For example, they reveal a degree of distinction between Ireland and Scotland/Wales, despite their close linkage in UEP-based analyses owing to their shared high frequency of the R1b3 haplogroup.

We next attempted to formally quantify Irish and Norwegian ancestry proportions in the NSG using six different admixture estimators, each of which considers
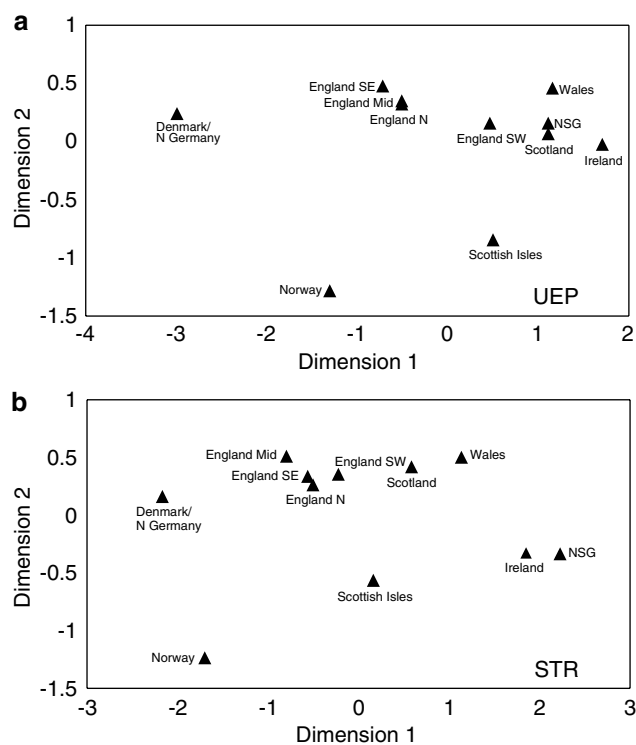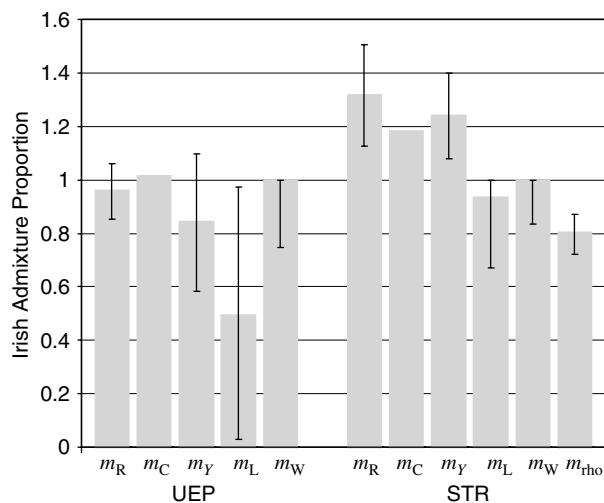

Figure 1 Y-chromosome haplogroup/haplotype frequencies in the Irish NSG and Irish and Norwegian potential parental populations. The modal STR profiles for specific lineages are (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393): AMH (14, 12, 24, 11, 13, 13); 2.47 (14, 14, 22, 10, 11, 13); 3.65 (16, 12, 25, 11, 11, 13). '+1' refers to one-step (STR repeat unit) mutational neighbours.

Figure 2 Inter-relationship of the NSG and other North-Atlantic European populations. The graphs plot dimension 1 *versus* dimension 2 of the variation in pairwise genetic distance matrices based on (**a**) UEP and (**b**) STR Y-chromosome markers. There is a clear separation in both cases, of the Irish and Norwegian parental populations. The closer affinity of the NSG to Ireland is also readily apparent, especially in STR variation. The S-stress and RSQ values for the MDS analysis in plot (**a**) are 0.038 and 0.995, respectively, whereas those for plot (**b**) are 0.039 and 0.994, respectively.

different factors and complexities in its execution. The results of this survey are shown in Figure 3. There is good correlation between most estimators with virtually all, whether using UEP or STR markers, supporting a substantial majority contribution from Ireland to the NSG.

The unconstrained estimators (m_R, m_C and m_Y) often returned notional Irish ancestry proportions greater than 1. Interpreting these as a 100% Irish contribution, the average estimate is ~90% Irish admixture with a standard deviation (SD) across estimates of 15.4%. The maximum likelihood m_L estimate based on UEP variation is exceptional in suggesting a minority (49%) Irish ancestry in the NSG. If this outlier is excluded the average estimate increases to ~95% and the SD more than halves to 7.3%. m_L calculation based on more numerous STR-defined alleles (haplotypes) returns an estimate of 93% Irish ancestry in agreement with other methods.

In some instances, a median value of m_L (reported in Figure 3) can give a poor representation of the results over the entire MCMC run.[27] The complete distribution of Irish ancestry estimates over the run of 50 000 MCMC iterations is shown in Figure 4. Analysis based on UEP-defined alleles (11 haplogroups) did not converge toward a consistent estimate whereas corresponding STR-defined alleles (106 haplotypes) show an unambiguous tendency toward a high Irish ancestry proportion estimate. Substantially longer chains (up to 250 000 MCMC iterations) were not found to aid UEP estimate convergence. It appears that the majority Norwegian ancestry m_L estimate is an artefact of poor LEA convergence.
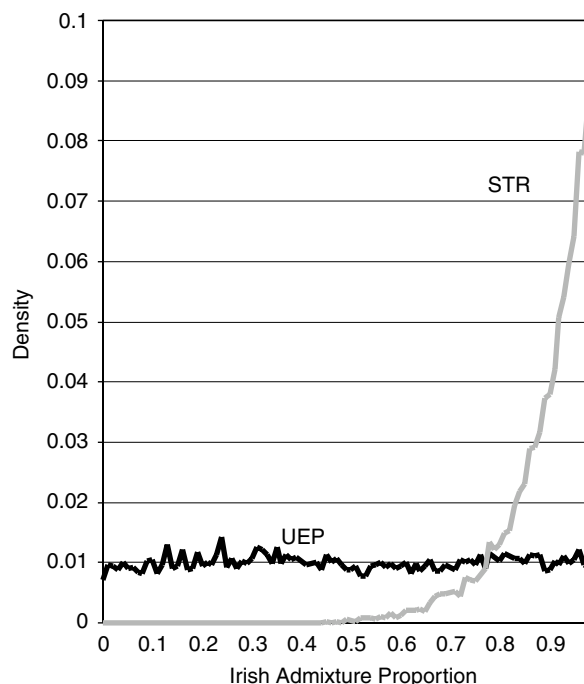
We also conducted additional admixture analysis with these methods using independent parental populations data sets including a larger Irish grouping ($n = 221$)[11] and separate Norwegian ($n = 250$),[6,30,31] Danish ($n = 69$)[6,32] or Swedish ($n = 110$)[30] populations. To allow comparability between data sets, this analysis was based on five STR haplotypes (DYS19, DYS390, DYS391, DYS392 and DYS393). However, all estimates, whatever the Scandinavian parental population, returned a ~90% Irish contribution or greater. Consequently, there is no evidence of major Scandinavian ancestry in the NSG from any of our analyses.

Discussion

There is little convincing evidence for substantial Scandinavian patrilineal ancestry in a cohort of Irish men bearing surnames of putative Norse origin. The absence of an appreciable legacy in both the general Irish population sample and this subgroup with a potential historical/cultural link to Norse society suggests a very limited general Norse genetic legacy in Ireland dating to the Viking period (ca. AD 800–1200). Substantial Irish ancestry in the NSG is not unexpected given historical evidence of

Figure 3 Irish (P_1) admixture proportion estimates for the NSG. Analysis was carried out using six different estimators (m_R, m_C, m_Y, m_L, m_W, m_ρ) independently for UEP and STR Y-chromosome diversity (m_ρ was developed for STR variation and was therefore not estimated for UEP haplogroups). Associated confidence boundaries for each estimator are also shown (with the exception of m_C as the LEADMIX software does not return a confidence interval for single-locus estimates). m_R, m_C and m_Y are unconstrained estimators and may return notional admixture proportions greater than 1.

Figure 4 MCMC convergence of the m_L admixture estimator. Each line shows the posterior probability density function of Irish (P_1) ancestry, over 50 000 iterations, based on UEP Y-chromosome variation (11 alleles) or six locus STR haplotypes (106 alleles). Whereas the UEP run showed little convergence, that based on STR variation showed an unambiguous tendency toward a high Irish ancestry proportion estimate.

assimilation between the native Irish and Norse settlers. However, the Norse genetic component of the distinctive hybrid Hiberno–Norse culture appears absent. It may be that the number of Norse and their descendents who settled and remained in Irish Norse colonies was small compared to their overall population; perhaps restricted to a thin, upper, stratum of these societies. The bulk of the Hiberno–Norse settlement population over the longer term may simply have been culturally adapted indigenous Irish.

Although the use of surname information to investigate historical questions is potentially powerful[11,16,17] any inference inherently relies on the accurate assessment of surname origin. The perturbed nature of Irish surname history means that similar or identical names can have more than one source.[12,17,33] For example, the name Sweetman can be of Norse origin or English toponymic from Swettenham.[12] The presence of several haplotypes within the NSG not found in the Irish or Norwegian samples may hint at this possibility. For example, there are multiple instances of the E3b haplogroup that is more typical of England than either Ireland or Scandinavia.[9] Alternatively, unique haplotypes, might indicate the action of drift over time. This could distort current allele frequencies and consequently population relationships and admixture estimates.

The proposed model of Viking migration is based not only on a Norse root for the surnames examined but also requires that the personal names from whence these derived were largely restricted to Hiberno–Norse settlements. There is some evidence that they had a degree of popularity in a native Irish context[34] and if such exchange was common it may mean that Norse names had begun to become disassociated with Norse colonial societies by the time of widespread surname adoption. The original personal names/nicknames could also have been introduced through a Norse female or indeed any other nonpatrilineal ancestor leaving no Y-chromosome trace when these names eventually became incorporated into hereditary surnames.

Finally, there is also an expected rate of disassociation of surname from initial ancestry owing to nonpaternity, adoptions and other causes. We have previously estimated this rate at 1.6% per generation,[17] which would be expected to roughly halve any initial Scandinavian ancestry fraction over $\sim$1000 years ($\sim$40 generations). The presence of multiple haplotypes within some surnames may be indicative of this, although it may also represent multiple surname foundations/origins.

The foregoing possibilities are not mutually exclusive and all may contribute to the overwhelming Irish ancestry observed in the putative NSG. Under a 'worst-case' scenario, where all putative Norse surnames have an alternate origin, the results apply to a general albeit small Irish population sample and still indicate very little Norse

migration to Ireland. However, the use of a variety of names and the inclusion of multiple bearers should help to cushion the impact of potential confounding factors (eg, some names not having a true Norse origin or the effect of subsequent nonpaternity events), allowing the identification of a grouping that is, at least, substantially enriched for Norse patrilineal ancestry.

Although open to interpretation, the limited Scandinavian admixture is a consistent and robust finding across several different admixture estimators, which encompass a range of considerations and computational frameworks. The choice of the most appropriate single estimator will depend on the expected relative importance of potential confounding factors. For example with recent admixture or UEP data, the effect of mutation is of less importance than drift. Despite the potential limitations of a relatively small number of alleles when using UEP haplogroups or the 'noise' from recurrent/back mutation associated with highly variable multi-STR haplotypes, both marker types generally return consistent and sensible results. UEP diversity is often favoured in population comparisons because of stability and obvious structure. However, STRs have potentially wider application in measuring diversity and finer-scale intra-haplogroup relationships. The use of a small number of STRs with moderate mutation rates that can be conveniently typed in a single multiplex may represent an under- utilised tool in macro-Y-chromosome population genetics.

The m_L maximum likelihood method[27] based on UEP Y-chromosome variation was a notable exception in our analysis as a median estimate indicated majority Scandinavian ancestry in the NSG. Although it performed well in STR-based analysis, where a large number of alleles are used, it appears that the method is considerably less accurate when the number of alleles is small. This observation may explain, in part at least, previous m_L Y-chromosome admixture estimates indicating high (35%) Scandinavian ancestry in a Dublin hinterland sample.[9] It may also be a relevant consideration in resolving the widely discrepant estimates of modern European ancestry attributed to the Neolithic or Palaeolithic periods using the same Y-chromosome data but different approaches, including m_L.[35,36]

Although the conclusions are predicated on the assumption that our cohort of Norse surnames largely originated in Hiberno–Norse society, they are consistent with limited Norse migration to Ireland. This observation may fit into a broader emerging picture, revealed by Y-chromosome and mitochondrial DNA, of Viking flexibility and adaptability in their settlement patterns across the North Atlantic. These included roughly equal male and female contributions through family-based migration, to secure settlements near Scandinavia and the integration of Gaelic females on frontier settlements further west, predominantly founded by Norse males.[8] In Ireland successful

settlement may have involved the wholesale incorporation of the indigenous Irish to create a distinctive, fusion Hiberno–Norse colonial society.

References

1 Davies N: *The Isles – A History*. London: Papermac, 2000.
2 Duffy S: *The Concise History of Ireland*. Dublin: Gill and Macmillan, 2000.
3 Ó Cróinín D: *Early Medieval Ireland 400–1200*. Essex: Longman, 1995.
4 Bradley J: The interpretation of Scandinavian Settlement in Ireland; in Bradley J (ed): *Settlement and Society in Medieval Ireland*. Kilkenny: Boethius Press, 1988.
5 Helgason A, Sigurethardottir S, Gulcher JR, Ward R, Stefansson K: mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 2000; **66**: 999–1016.
6 Helgason A, Sigurethardottir S, Nicholson J *et al*: Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 2000; **67**: 697–717.
7 Helgason A, Hickey E, Goodacre S *et al*: mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001; **68**: 723–737.
8 Goodacre S, Helgason A, Nicholson J *et al*: Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. *Heredity* 2005; **95**: 129–135.
9 Capelli C, Redhead N, Abernethy JK *et al*: A Y chromosome census of the British Isles. *Curr Biol* 2003; **13**: 979–984.
10 Jobling MA: In the name of the father: surnames and genetics. *Trends Genet* 2001; **17**: 353–357.
11 Hill EW, Jobling MA, Bradley DG: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
12 MacLysaght E: *Irish Families: Their Names, Arms and Origins*. Dublin: Irish Academic Press, 1985.
13 Bosch E, Lee AC, Calafell F *et al*: High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci Int* 2002; **125**: 42–51.
14 Y Chromosome Consortium: A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
15 Jobling MA, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003; **4**: 598–612.
16 Moore LT, McEvoy B, Cape E, Simms K, Bradley DG: A Y-chromosome signature of hegemony in Gaelic Ireland. *Am J Hum Genet* 2006; **78**: 334–338.
17 McEvoy B, Bradley DG: Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet* 2006; **119**: 212–219.
18 Choisy M, Franck P, Cornuet JM: Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Mol Ecol* 2004; **13**: 955–968.
19 Roberts DF, Hiorns RW: Methods of analysis of the genetic composition of a hybrid population. *Hum Biol* 1965; **37**: 38–43.
20 Bertorelle G, Excoffier L: Inferring admixture proportions from molecular data. *Mol Biol Evol* 1998; **15**: 1298–1311.
21 Dupanloup I, Bertorelle G: Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* 2001; **18**: 672–675.
22 Long JC: The genetic structure of admixed populations. *Genetics* 1991; **127**: 417–428.
23 Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE: Caucasian genes in American blacks: new data. *Am J Hum Genet* 1992; **50**: 145–155.
24 Wang J: Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* 2003; **164**: 747–765.
25 Nachman MW, Crowell SL: Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000; **156**: 297–304.
26 Zhivotovsky LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
27 Chikhi L, Bruford MW, Beaumont MA: Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* 2001; **158**: 1347–1362.
28 Schneider S, Roessli D, Excoffier L: ARLEQUIN Version 2.000: software for population genetic analysis. Genetics and Biometric Laboratory, University of Geneva, 2000.
29 Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB: Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 2001; **98**: 5078–5083.
30 Zerjal T, Beckman L, Beckman G *et al*: Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* 2001; **18**: 1077–1087.
31 Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG: Y chromosome evidence for Anglo-Saxon mass migration. *Mol Biol Evol* 2002; **19**: 1008–1021.
32 Bosch E, Rosser ZH, Norby S, Lynnerup N, Jobling MA: Y-chromosomal STR haplotypes in Inuit and Danish population samples. *Forensic Sci Int* 2003; **132**: 228–232.
33 MacLysaght E: *The Surnames of Ireland*. Dublin: Irish Academic Press, 1985.
34 Ó Cuív B: Personal names as an indicator of relations between native Irish and settlers in the Viking period; in Bradley J (ed): *Settlement and Society in Medieval Ireland*. Kilkenny: Boethius Press, 1988.
35 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic *Homo sapiens* sapiens in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
36 Chikhi L, Nichols RA, Barbujani G, Beaumont MA: Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 2002; **99**: 11008–11013.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)