doi:10.1038/nature14230

The fine-scale genetic structure of the British population

In this Supplementary Note we provide further historical background, expand on some of the issues raised in the Discussion section of the main text, and provide further comments on our analyses. High-resolution PDF files of the extended data figures are available at (http://www.well.ox.ac.uk/POBI).

Archaeological, linguistic, and documentary evidence for the peopling of the British Isles

We briefly summarise the major population groups and movements of people within and into the UK, based on archaeological, historical and linguistic evidence (see Cunliffe (2012)[1] for further background detail).

Although Britain was populated prior to the last glaciation, no permanent human settlement survived the glacial period. Palaeolithic hunter-gatherers were able to colonise the British peninsula as the climate warmed at the end of this glaciation (9600BC) because sea levels were still low and Britain was joined to Continental Europe by land across the Southern North Sea (Fig. 3a). Ireland had already become separated by this date and was colonised around 8000BC in the Mesolithic, by hunter-gatherers moving by boat perhaps from western parts of Britain and along the Atlantic coast of Europe, a sea route also likely to have been used by settlers into the western parts of Britain.

Britain became separated from mainland Europe by rising sea levels around 6500BC[2]. Archaeological evidence suggests that good communications with continental Europe were maintained, though the existence or extent of any possible migrations is not known. Agriculture reached the British Isles at the start of the Neolithic (4000BC) and cereal cultivation is thought to have spread throughout Britain and Ireland within only 150 years[3]. Distinctive beaker pottery spread into Britain (2500BC) shortly before the start of the Bronze Age (Extended Data Fig. 7a). The styles of beakers establish that there was both a western (Atlantic coast) and eastern (Southern North Sea/Eastern Channel) route from the Continent. Both the start of the Neolithic and the Beaker Period have been argued as times of major migration into Britain.

The Iron Age (from 800BC) was a period when very distinctive regional cultures developed in Britain with identity being reflected in pottery styles[1], probably implying some limitation of population movement in Britain. At the time of the Roman conquest of Britain (43), there were well-established Iron Age tribal groupings across the UK (shown schematically in Extended Data Fig. 7b). This is the earliest period for which there is documentary evidence of the boundaries of various groupings, although "tribal" organisation is also likely to have been a feature during at least some earlier periods.

Roman control of Britain (43-410) extended north to Hadrian's Wall and sometimes beyond, but was exercised differently in different regions (Fig. 3b). The most heavily Romanised area of Britain was the south east where farming was re-organised to include the villa system. Here the resulting market economy fostered the development of numerous small towns and there was a Roman civilian administration. Elsewhere, the local populations were allowed to maintain their

traditional ways. During the Roman period there was some movement of people into the UK from other parts of the Roman Empire, especially soldiers from Gaul and later Germany, though this amounted to at most a few per cent of the total population[1].

There was large-scale settlement by Angles, Saxons, and possibly Jutes and Frisians (collectively known as Saxons) from the Danish peninsula and the north west German coast into Southern and Eastern Britain especially during the period from 450-500. This followed the end of Roman rule and was facilitated by the ensuing collapse of social systems and population numbers (Fig. 3c). The Saxon migration involved changes in language (from Brythonic Celtic to Old English), place names, material culture, and even cereal crops. However, Wales, south west Scotland and initially south west and north west England remained under the control of Brythonic Celtic-speaking Britons. During this period there was also movement of the Scots (originally a northern Irish people) between Ireland and the west coast of what is now Scotland within the Goidelic Celtic-speaking kingdom of Dalriada.

Vikings from Norway settled in Orkney and other islands off the north of Scotland, along the north coast of Scotland, and in the Western Isles (off the west coast of mainland Scotland) from the late 8th Century (Fig. 3d). Norway annexed Orkney as an earldom (875-1468) and Orkney's culture became entirely Scandinavian. There is also evidence of Norse Viking settlements, on a smaller scale, in Ireland and in Wales. Viking raids on England began in the late 8th century and, following a large-scale invasion in 865, Danes began to settle a swathe of land from East Anglia to north west England, which became the Danelaw (Fig. 3d). Many Scandinavian place names from this region have survived to the present day but Scandinavian material culture was soon lost and it is not thought the scale of settlement was large.

The Norman invasion of England in 1066 resulted in a small Norman elite establishing control over all of England, South Wales and the east of Ireland, but relatively little population movement into the UK. The Normans were based in northern France, with ancestry from the Danes, Franks and Bretons.

In the 400 years leading up to 1468, the Scots of Dalriada, the Picts of north east Scotland, the Britons of Strathclyde and Galloway, the Saxons (Northumbrians) of Lothian, the Gaelo-Norse from the Western Isles to Kintyre and the Norwegians of Caithness, Orkney and Shetland were brought together in the kingdom of Scotland.

Two further events were relevant to the peopling of rural areas of the British Isles. English and Flemish people were settled in Pembrokeshire (south west Wales) during the 1100s, and settlers for the "Ulster Plantations" (migrations into what is now Northern Ireland) of the 1600s were recruited from south west Scotland and Northern England.

Estimates of population size in the UK throughout prehistoric and early historical times are necessarily imprecise. The population of the British Isles in 9000BC has been estimated at ~1,100 Mesolithic hunter-gatherers[4]. The population is then thought to have increased to 2,750-5,500 by 5000-4000BC, towards the end of the Mesolithic[4]. The introduction of agriculture at the start of the Neolithic (around 4000BC) greatly increased the size of the population that could be supported in the British Isles. The early Neolithic (~3000BC) population of Ireland has been estimated as 40,000 and of Britain as 100,000 with the combined population of Britain and Ireland reaching 500,000 by 1000 BC, towards the end of the Bronze Age[5]. By the Roman Conquest of 43, the Iron Age population of what was to become

Roman Britain has been estimated as 1.5-2 million increasing to 2-5 million during Roman control[6]. The collapse of Roman rule has been suggested to have led to a significant population decline, by a factor of two or more[6], within the formerly Romanised region. Populations were rising again from the start of the middle Saxon period (600) and by 1068 (Domesday) the population of England was 1.4-1.9 million[7].

fineSTRUCTURE Analyses

Here we focus on issues relating to the initial fineSTRUCTURE clustering analyses as depicted in Fig. 1 and SI Fig. 1. In particular, we extend the discussion of some of the notable features of the clustering analyses; report on the robustness of our fineSTRUCTURE analysis; make some further observations about the possible relationships between the observed clusters and known historical and demographic events; and show further applications of the genetic analyses.

Notable Features of the fineSTRUCTURE Clustering Analyses

SI Figs. 1.1 to 1.24 show the equivalent plots to those shown in Fig. 1 for all levels of the hierarchical clustering tree from 2 clusters to 24 clusters and then 53 clusters. One can 'step through' these figures, starting from the coarsest clustering of the samples into two groups, and see the finer levels of structure emerge. Here we focus on the UK clusters inferred in our fineSTRUCTURE analysis (those subplots labelled 'a' in SI Fig. 1). In the main text we described the major features of the splitting of the UK sample into 17 clusters. Here we point out some other interesting features that emerge as one examines finer and finer splits. At 18 clusters further differentiation is observed in Orkney, which is also well localized. At 21 clusters the Welsh borders cluster splits into two parts, one in the north and one in the south. At 23 clusters we observe a tight and distinct cluster at the tip of Cornwall.

Even for UK clusters which are well localised geographically, we typically observe some level of sample overlap with neighbouring clusters. This is to be expected for a number of reasons. Firstly, most clusters do not have hard geographical barriers. Secondly, samples are indicated on the map at the centroid of the birthplace of their four grandparents. If, for example, an individual had three grandparents from within one cluster, and a fourth from another region, they may well still be assigned to the cluster associated with the majority of their grandparents, but their location on the map will be moved towards the fourth grandparent, possibly outside the bulk of the cluster to which they are assigned.

It is instructive to consider the information contained in the measure of confidence we defined and used for the assignment of individuals to clusters (see Methods). Recall that for each individual i, $P_{J,i}$ is a J-vector, with one component for each cluster at a given level L_J. Each component is the measure of the confidence of the assignment of the individual i to each cluster, and the maximum of these gives the cluster assignment for individual i. For simplicity call this maximal value m_i.

Extended Data Fig. 1 illustrates the distribution of m_i across individuals in a very particular way, allowing one assessment of the information contained in the measures. In this case, focussing on the level of the hierarchical clustering tree containing 17 clusters, we set a threshold $t = 0.7$ (chosen for illustrative purposes only) and observe that the overwhelming majority of assignments are 'confident' (i.e. $m_i > t$). When individuals do have uncertain assignments (i.e. $m_i \leq t$) they are

usually located in regions where two clusters adjoin or overlap each other, as one might expect. In particular we observe uncertain assignments in *Northumbria*, *Cumbria*, and *W Yorkshire*, and where these clusters overlap the *Cent./S England* cluster, as well as on the border of *Cornwall* and *Devon*. This leads us to believe that the measures we have defined are useful reflections of the confidence that an individual is assigned to a particular cluster.

We observe similar patterns for other thresholds and at other levels of the hierarchical clustering tree (data not shown).

There are instances of individuals assigned to a particular cluster who are located a considerable distance from the other members of the cluster. For this to be the case, the individual's grandparents must be born near the map location where the individual is represented, and the grandparents (or presumably at least three grandparents) must be genetically representative of the cluster to which the individual is assigned. This could occur if a whole community moved within Britain and subsequently preferentially married within their community. Documented examples of this occurred when mining communities from south-west England were incentivised to move elsewhere in the country by mine-owners eager to bypass a striking local workforce; when the incoming workers were ostracized by the locals, they tended to live principally within the community of migrants.

Robustness of the Clusters

We assessed convergence of the fineSTRUCTURE MCMC runs in various ways. This included running independent chains, and comparing aspects of the assignments of individuals to clusters, and the results of downstream analyses, between the two chains. Reassuringly, given the size of the state space being explored, these diagnostics confirmed mixing of the MCMC chains, and their convergence. See Extended Data Fig. 2 for diagnostic plots for both the UK and European fineSTRUCTURE clustering analyses.

A potential concern with our UK clustering analyses is that they may be capturing excess recent relatedness induced by our sampling scheme, rather than real underlying population structure. Our QC procedures exclude one individual from any pair of close relatives. This is done on the basis of a pairwise identity by descent (IBD) statistic[8], where we exclude one member of any pair of individuals with relatedness greater than 0.05. Thus any potential concern would be with more distant relatives. If our analyses are capturing excess recent relatedness then some pairs of individuals within the inferred clusters should share greater portions of their genome IBD. To address this possibility we compare the distribution of the pairwise IBD statistic within clusters to that of the IBD statistic across the whole sample. Extended Data Fig. 4 depicts the distribution of the pairwise IBD statistic both within the inferred clusters and across the whole UK sample for the level of the hierarchical tree we focus on in our main analyses (17 clusters), and the finest clustering of our sample (53 clusters, see SI Fig. 1.24).

For the clusters we focus on in our main analyses, it is clear that for the most part the distribution of the IBD statistic matches the distribution across the whole of the UK. The exceptions are two of the clusters in Orkney (*Westray* and *Orkney 2*), *NE Scotland 1* and the *N Pembrokeshire* cluster. In these cases the distribution is centred considerably lower than that of the whole sample, indicating that these clusters contain samples that share more of their genome IBD than is typical for arbitrary pairs of individuals in our study. In the case of the Orkney clusters this is

readily explained by the small population size and relative isolation of those islands. Population structure is actually caused by shared ancestry, though typically not over very recent timescales, so that we might expect some slight increase in relatedness within real genetic clusters. The key observation is that for the most part there is no evidence that groups of close relatives were over-represented in our sampling scheme, or that this will have affected our downstream analyses. At the finest level of the hierarchical clustering tree one observes that in general the small clusters share more of their genome IBD (as measured by our statistic) than is typical for the UK. This is to be expected as fineSTRUCTURE will rightly detect small groups that share more of their genomes IBD. Crucially, for all clusters containing ten or more individuals, except for those in Orkney, the distribution of the IBD statistic is well-matched to that across the whole sample and, as stated above, the case of Orkney is readily understood. We thus conclude that the fine-scale population structure we observe is not an artefact of our sampling scheme.

As noted in the main text and Methods, fineSTRUCTURE characterises each cluster by a 'copying vector', which summarises, for that cluster, the proportion of its closest ancestry that comes from individuals across each of the clusters. In fact, this copying vector can be calculated for any group of samples. One can use these vectors to test if the clusters inferred by fineSTRUCTURE are capturing significant differences in ancestry, and to give a sense of the strength of the differences observed. Given a pair of inferred clusters and their copying vectors one can calculate the total variation distance (TVD_{CV}) between the pair. Furthermore, for this pair of clusters one can randomly reassign the individuals in the clusters, maintaining the cluster sizes, and then recalculate the copying vectors and the total variation distance between them. Repeating this process one can obtain a p-value from a permutation test of the null hypothesis that, given the cluster sizes, the individuals in the two clusters are assigned randomly to each cluster (see also Methods). Supplementary Table 3 shows the value of the TVD_{CV} statistic for all pairs of the 17 clusters used in our main analyses. Each of these is compared with a null distribution based on 1,000 permutations and a p-value is calculated. All the pairwise comparisons of clusters give p-values below 0.001, confirming that the clusters inferred by fineSTRUCTURE are capturing highly significant ancestry differences.

The ancestry differences between the clusters we infer are significant. Nonetheless, the population structure is subtle. Estimated F_{ST} values between the clusters represented in Fig 1 are small (average 0.002, maximum 0.007, Supplementary Table 2) which, although larger than that between the sampling locations, is still indicative of very limited population structure.

Relationship of the Clusters to Known Events

The similarity between the genetic clusters in Fig. 1, and the geo-political boundaries in Fig. 3c (600, after the major Saxon migrations) is noteworthy. Regions of Britain outside those most directly controlled by the Romans maintained much of their local identity, even under Roman rule. These may well have resumed their tribal identities with the collapse of Roman control, in turn maintaining some degree of isolation from neighbouring groups. Most were not directly affected by the large-scale Saxon migrations from 450-500, and only came under Saxon control much later, if at all. Comparing Figs. 1 and 3c shows UK genetic clusters located in roughly the region of the kingdoms of Rheged (*Cumbria*, white triangles), Elmet (*W*

Yorkshire, blue triangles), Dalriada (*N Ire./W Scotland*, light green triangles), Gwynedd (*N Wales*, green squares), Dyfed (*N Pembrokeshire*, pink squares and *S Pembrokeshire*, yellow inverted triangles), and Dumnonia (two groups: *Cornwall*, pink crosses, and *Devon*, blue circles, see Fig. 1). Following the expansions into Scotland from the kingdom of Dalriada in the west, and the Saxons from the south, the Picts were restricted to the northeast[9], although the numerical scale of this movement is not known. Nevertheless, this movement of Picts might possibly be reflected in one or both of the two UK clusters observed in northeast Scotland (*NE Scotland 1*, white squares and *NE Scotland 2*, pink circles). As Fig. 3c illustrates, there was a linguistic barrier between the Saxon regions and the rest of the UK, where various Celtic languages were spoken, some of which still survive.

It is also noteworthy that the large *Cent./S England* cluster (red squares) largely coincides with the region of the UK under most direct Roman control (Fig. 3b), and is close to the region under Saxon control in 600 (Fig. 3c). Plausibly the effect of Roman control was to break down the Iron Age tribal entities in the region (Extended Data Fig. 7b), and hence to reduce geo-political barriers to movement, as well as facilitating movement through improved roads, and encouraging it through limited urbanisation (which declined after the Roman period). Saxon control of roughly the same area, although at times divided into several large kingdoms, did not reintroduce many geo-political barriers to movement.

There are several examples in Fig. 1 of clusters occupying the same geographical area, including in Northern Ireland (*N Ire./S Scotland*, *N Ire./W Scotland*) and northern England (*Cumbria*, *Northumbria*, and *N Ire./S Scotland*). Genetic clusters in the same area will lose their distinctiveness over time through intermarriage, unless mating occurs largely or exclusively within clusters. This could occur for human populations if there are linguistic, religious, or other cultural barriers between the groups. This may well account for the overlapping clusters in Northern Ireland. Soon after 1600, following the British conquest of Ulster, there was an organised, extensive, migration of people from Scotland and northern England into six of the eight counties of Ulster (which became modern Northern Ireland), in what is known as the "Plantation of Ulster". The size of this migrant population has been estimated at up to 80,000 by the 1630's. These were almost all English speaking Protestants who outnumbered the Gaelic-speaking Catholic indigenous population[10]. The *N Ire./S Scotland* cluster most probably reflects descendants (on both sides of the Irish Sea) of this historical population movement. On the other hand, it may be that the distinctive clusters observed in northern England may represent a transient phenomenon where groups which were previously distinct genetically and geographically have migrated beyond their original boundaries and are in the process of admixing.

Maps and Visualization

On a practical level our genetic clustering was robust enough to allow us to identify and correct some individuals who had been geocoded incorrectly. As noted in the Methods, the latitude and longitude for each UK sample's grandparents' birthplaces was assigned automatically using a place name gazetteer. In some cases the genetic clustering was used to check the accuracy of the automatic geocoding, and identify samples that seemed geographically separated from their genetic cluster. This enabled the identification of several samples that had been geocoded incorrectly, which was confirmed by checking the original documentation for the sample

collection. For example, a number of samples were automatically, and erroneously, geocoded to Blackburn, Lancashire when in fact the project records showed they should have been geocoded to Blackburn, Aberdeenshire. For all samples the geocoding was checked manually to exclude typographical errors and errors in the identification of place names.

Comparison to Other Methods (PCA and ADMIXTURE)

We applied two other methods commonly used for detecting population structure to our data – principal components analysis (PCA) and the program ADMIXTURE (see Methods). The results for PCA are shown in SI Figs. 1.1 to 1.24 (panel **c**) and Extended Data Fig. 3a. In the former case the samples are plotted against the first two principal components using the symbol of the cluster to which they are assigned by fineSTRUCTURE for the given level of the hierarchical clustering tree. In the latter case plots for all pairs of the first five principal components are shown, with the samples coloured to indicate the collection district from which they were taken.

As with previous PCA analyses of UK data (e.g. [11]), there is a roughly north-south cline from the top left towards the middle right of the plot of the first two principal components (visible with samples either coloured by sampling region as in Extended Data Fig. 3a or according to our fineSTRUCTURE clustering, SI Fig. 1), with the Welsh samples separated from the middle right of the plot towards the bottom left. While PCA thus broadly separates samples from Orkney, and separately most samples from Wales, it does not resolve anything beyond the first few splits in the tree of our primary analyses (and these not perfectly), much less the fine-scale distinctions in our analyses, even with the inclusion of additional principal components (Extended Data Fig. 3a).

ADMIXTURE[12] is a commonly used program to infer clusters or subpopulations of individuals on the basis of genetic data. Unlike fineSTRUCTURE, when applying ADMIXTURE one sets K, the number of clusters into which the samples are to be divided, in advance as a fixed parameter of the model. (There is a method for choosing the 'best' value of K using cross-validation, but we restricted ourselves to the most straightforward analysis.) We ran ADMIXTURE three times on our data, to test the ability of the algorithm to detect structure on both the fine- and coarse-scale. For the coarse-scale we set K=2 and K=3. For the fine-scale we set K=17, enabling comparison with our main analysis using fineSTRUCTURE. The results are shown in Extended Data Fig. 3b. When K=2, so that ADMIXTURE divides the UK sample into two groups, one group contains almost all the samples in Orkney, but also some samples from Wales, Scotland and Northern Ireland, with the other group containing the remaining samples (predominantly the samples from the UK but not Orkney). The separation of Orkney from the rest of the UK in this case is slightly less precise than that obtained using fineSTRUCTURE. When K=3, the three clusters found by ADMIXTURE are: a cluster which contains virtually every Orkney sample and very few others; a cluster that contains almost all of the samples from Wales and highland Scotland, but also some but not all samples from Northern Ireland, south west Scotland, NE Scotland and a scattering of English samples; and a cluster which contains almost all of the English samples plus some from elsewhere. For both K=2 and K=3 the structure inferred by ADMIXTURE is consistent with both geography and the linguistic and historical record, although the clustering and these relationships are less clear than that obtained for three clusters using

fineSTRUCTURE. When K=17, ADMIXTURE fails to find any fine-scale population structure except for clusters representing Orkney and Wales. Interestingly, as with applying fineSTRUCTURE, one observes a split in Orkney between the southern and northern islands. However, for the most part the clusters are not localised geographically and are not straightforward to interpret.

Ancestry Profiles

We now turn to a fuller discussion of the ancestry profile analyses, the results of which are given in Fig. 2, Extended Data Fig. 6 and Supplementary Table 4.

Information about the relative ordering of some migrations

A critical observation in the main text is that groups which contribute significantly to the ancestry profiles of all UK clusters most probably represent, at least in part, migration events into the UK that are relatively old, since their DNA had time to spread throughout the UK. Conversely, groups that contribute to the ancestry profiles of only some UK clusters most probably represent more recent migration events, with the resulting DNA not yet spread throughout the UK by internal migration. If DNA from these latter groups had reached all the UK clusters, the pattern of ancestry we observe would involve the independent loss, in different regions of the UK, of several different ancestry contributions from widely separated parts of Europe, and we regard this as unlikely.

As noted in the main text, "old" and "recent" are relative terms – we can infer the order of some events in this way but not their absolute times. Also, although we refer to "migration events" we cannot distinguish between movements of reasonable numbers of people over a short time or on-going movements of smaller numbers over longer periods of time. The "old" migration events to which we refer here and below represent the earliest migration events from which substantial DNA survives to the present. This may reflect the peopling of Britain after the last ice age, but could also represent subsequent migrations if these effectively replaced existing populations.

England and Wales

Examination of the ancestry profiles of the various UK clusters (visible in the columns of the bar chart in Fig. 2) reveals some interesting shared patterns. One distinct overall pattern appears in the ancestry profiles of three of the UK clusters (from north to south: *N Wales*; *N Pembrokeshire*; *S Pembrokeshire*): absence of GER3 and FRA17, presence of FRA12, and relatively higher proportions of GER6 and FRA14. Interestingly, these are the three clusters located in Wales. A second general pattern is shared by a number of other UK clusters (from north to south: *Northumbria*; *Cumbria*; *W Yorkshire*, *Cent./S England*; *Welsh Borders*; *Devon*; *Cornwall*): significant presence of GER3, absence of FRA12, relatively higher contributions from groups FRA17 and DEN18, and relatively lower contributions from FRA14. These are the seven UK clusters located within England. These shared patterns across the ancestry profiles of different UK clusters also emerge from a correlation analysis of the data (see Supplementary Table 7).

Norway and Sweden

We see a significant contribution to the ancestry profiles of some UK clusters from groups in Norway (Fig. 2: groups NOR53-NOR90, pink through to purple). This contribution is largest for the three clusters from Orkney (Fig. 2, *Westray*, *Orkney 1*, *Orkney 2*), where it totals 24%, 24% and 20% respectively. The next largest

contribution is to *N Ire./W Scotland* (17% total, Fig. 2), with declining contributions moving south through Scotland (10-11%) and England (3-7%), and also some contribution in Wales (*N Wales*, *N Pembrokeshire*, *S Pembrokeshire*: 7%, 5%, 5%). Our genome-wide analyses are thus qualitatively consistent with earlier genetic analyses of single-marker systems[13–16] and consistent with the known historical migrations of Norse settlers (see above).

We observed considerable fine-scale population structure in modern-day Norway, with good geographical localisation of the different genetic groups (Fig. 2, Extended Data Fig. 5b). This potentially allows localisation of the Norwegian groups which contribute ancestry to the UK. Interestingly, many Norwegian groups, with quite varied geographical locations, contribute to the ancestry profiles in Orkney (and elsewhere in the UK). The largest contributions come from groups NOR53 (northern coast), NOR64 (around Oslo) and NOR90 (south-western coast, closest to Orkney), with little or no contribution from other groups in the south or on the southern coast. The simplest explanation for our observation is that several geographical regions of Norway contributed settlers, via the Norse Vikings, to Orkney. Other explanations are possible, although we believe considerably less likely. One possibility is that settlers to Orkney originated from one region in Norway, which was not directly sampled in our analysis, and that individuals from that region, or their ancestors, also migrated to distinct areas in Norway, making a combination of those regions the best contributors to the Norse part of the ancestry profiles of people in Orkney. We have reasonably extensive sampling within much of Norway, and in particular in the southern and western coastal regions historically associated with the Norse Vikings, so we think this explanation less likely. A related possibility is that the population in Norway at the time of the Vikings was much more homogeneous than it is now, with much of the structure we observe arising after that time, so that the Norse part of the ancestry profile of Orkney is best described as a mixture of several current groups in Norway, all descended from the source population for the Norse settlers to Orkney. Given the extensive physical barriers to movement in Norway, we believe it unlikely that the Norwegian population around the year 900 exhibited substantially less population structure than does the current population.

The bootstrap confidence intervals (CIs) in Extended Data Fig. 6a (and Supplementary Table 4) suggest a non-zero total contribution to all UK clusters (and in particular those outside Orkney, Scotland, and Wales) from Norwegian groups, and separately from Swedish groups. There are other sources of error in this analysis not captured by the bootstrap CIs, so we would advise caution in the interpretation of such low levels of contribution. If they are real, there are several possible explanations. Perhaps the most plausible is that they represent DNA which moved into the UK at an early stage from a population or populations elsewhere in Europe at least some of which also moved into Scandinavia, with haplotypes from that early ancestral population surviving in modern Britain and in modern Norway and/or Sweden. Other possibilities include: direct migration before the Viking era from Norway and Sweden, either as part of the early migrations into the UK (to allow the DNA to spread throughout the UK) or in a series of migrations to different parts of the UK; more recent migration from other regions of Europe which share ancestry with the UK into Scandinavia; or migration from the UK to Norway and Sweden. We note that analogous explanations could apply to other low level contributions found in the ancestry profiles of all or nearly all clusters. For

example, these could provide alternative explanations for some of the Danish contribution to all the UK clusters, although because of Denmark's proximity to the land bridge, we prefer the explanation advanced below of direct early migration from Denmark into the UK.

Earliest Migrations

Because they contribute substantially to the ancestry profiles of all of the UK clusters, we suggest in the main text that groups GER6, BEL11, and FRA14 all represent descendants of early migrations into the UK. Here we expand on that discussion. Group FRA14 is observed almost entirely in the sampling location of Rennes in north-west France, where the associated hospital has a large catchment area including Brittany, Normandy, and the Loire regions and could represent descendants of peoples from one or several of these regions (or, as always, from other regions whose descendants moved to the current sampling locations). Group GER6 is most prevalent in the west of Germany near the German-Netherlands border. Group BEL11 is one of the two groups in Belgium. The other Belgian group, BEL7, makes little or no contribution to the UK ancestry profiles. All of the Belgian sampling locations in our study are in Flanders, the more northerly part of Belgium, which borders the Netherlands. Samples from the Netherlands were not available for this study. In interpreting our results, it should be borne in mind that European groups (such as GER6 and BEL11) which contribute to the UK ancestry profiles could do so because they represent the best surrogates in our dataset for groups which we have not directly sampled (such as the Netherlands), in addition to, or instead of, more direct contributions. Better resolution of the origins of these differences will depend on finer sampling of the relevant European populations, as we have done in the UK. Archaeological evidence suggests two different routes for early migrations into the UK, one via the land bridge from Europe and another by sea from the Atlantic coast of Europe to Wales and other western parts of the UK. Given the current locations of these groups, and their contributions relative to each other to UK clusters, the simplest explanation consistent with this archaeological evidence is that the group currently close to the west coast of France (FRA14, contributing relatively more than the other two groups to the clusters in Wales for example) represents descendants of the sea-based migrations, whereas GER6 and BEL11, located near the region that once connected to the Doggerland land bridge, and contributing relatively more than FRA14 to the clusters in Engand, represent descendants of the early land-based migrations.

Saxons and Danes

The group GER3 makes no contribution to the ancestry profiles of the three UK clusters in Wales, nor to the cluster spanning Northern Ireland and western Scotland, from which we concluded that it likely represents a more recent migration to the UK. It, however, makes non-zero contributions to all seven of the UK clusters in England, though the contributions to the northwest and northeast of England are small. GER3 is localised in northern and north-western Germany, in the region from which it is known that many of the Saxon migrants to the UK originated. Furthermore, contributions from GER3 are higher in the parts of the UK known to have been settled by Saxons. This leads us to conclude that contributions from GER3 to the ancestry profiles of the UK clusters most probably result directly from the Saxon migrations. As we noted above, throughout, we use "Saxon" as shorthand for the Angles, Saxons, Jutes, and possibly Frisians, but in this paragraph it would equally apply in the stricter sense of the Saxon component of this larger grouping.

To make these distinctions would, again, need a much finer analysis of the relevant European populations.

Group DEN18, from modern Denmark, contributes to the ancestry profiles of all the UK clusters, although for some clusters at low levels. Migrants from Denmark could have entered the UK at many different times: in early migrations, either overland or later by sea; with the Saxon migrations (the Jutes, entirely, and the Angles, partially, originated in what is now Denmark); with the Danish Viking settlement; and potentially even with the Norman invasion, as Normandy was itself settled by Danish Vikings 100-200 years before the Norman invasion of the UK. The fact that contributions from GER3 are absent from Wales suggests that Saxon DNA (from GER3) failed to reach Wales in appreciable quantities by internal migration within the UK since the Saxon migrations. Consequently, some DNA best represented by that in modern Denmark must have reached the UK in early migrations, before the Saxon invasions and Viking era, for it to contribute to the ancestry profiles of all the UK clusters.

There are broad similarities between the pattern of contribution to the UK of DEN18 and GER3, especially if some part of the contribution of DEN18 is attributed to early migrations. This, and the geographical overlap of the modern Danish group with the locations of the Angles and Jutes, leads us to the tentative conclusion that a considerable proportion of the contribution of DEN18 also reflects the Saxon migrations.

Definitively separating Saxon and Danish Viking inputs is impossible, but we offer some insights. Danelaw, the area of England controlled by the Danish Vikings, was geographically limited (Fig. 3d), and there is no record of Danish Viking settlement in the southern areas of the large UK cluster in central and southern England (Fig. 1, red squares). In contrast, the Saxon migrations are known to have enjoyed a larger geographical spread (Fig. 3c) with much of what became England being part of Saxon kingdoms at the time of the Danish Viking invasion. Since our approach is powered to detect quite subtle levels of population structure, it is informative that we see no remnant of the Danelaw, in terms of a distinct genetic cluster within the UK. The greater the input of DNA from Danish Viking settlement, the greater the level of migration needed to produce the observed genetic homogeneity across central and southern England. We thus think it likely that there was limited input of DNA from Danish Viking settlement, and that the majority of any more recent ancestry contribution from Denmark reflects the Saxon migrations. As noted above, the Norman invasion in 1066 involved a small ruling elite, with limited influx of DNA.

Migrations from France

We argued in the main text that there is evidence for later migrations of people into Britain after the repopulation subsequent to the last ice age, but before any of the migrations known from historical records. These are best captured by FRA17 outside Wales, with migrations represented by FRA12 essentially only into Wales and Northern Ireland and/or Scotland.

The group FRA12 is essentially only present in Wales and the two clusters spanning Northern Ireland and Scotland, while FRA17 is absent from Wales. Although the individual confidence intervals (Extended Data Fig. 6a and Supplementary Table 4) for FRA12 contributions do not exclude zero for many of these clusters, and those for FRA17 do not exclude non-zero values, the pattern of non-zero FRA12 point

estimates, in the Welsh/Scottish and Northern Irish clusters, and zero FRA17 point estimates in Wales, is striking, and, we believe, informative. Interpretation is somewhat difficult as both FRA12 and FRA17 occur at all three of our main sampling locations in modern France. FRA17 is relatively more common than FRA12 in the north and northwest sampling locations, while FRA12 relatively more common in the central French sampling location, and this could account for their complementary contributions, especially to Wales. More precise geographical sampling in France would be needed to confirm this possibility.

Turning in more detail to the group FRA17, we note that it is one of the largest contributing groups to the ancestry profiles of the UK clusters. Its presence/absence pattern (notably its absence from Wales) strongly suggests that it results from a migration or migrations later than those of the earliest migrations which contributed DNA to the modern UK population (GER6, BEL11, FRA14): internal migration has spread the DNA from these early immigrants across the UK, so that if the migrations represented by FRA17 were earlier than or contemporaneous with these, then the same migrations should also have spread the resulting (FRA17-like) DNA throughout the UK, including Wales.

We also argue that the FRA17 contribution is unlikely to reflect any of the known movements of people in historical time (i.e. since the Roman invasion of Britain). The influx of people to Britain during Roman control is known to be small relative to the then population, and much too small to explain such a large contribution to the ancestry of many UK clusters. Next, as noted in Methods, it seems unlikely to result from the Saxon migrations. (For completeness we repeat those arguments again here.) These migrations did not directly involve people from what is now France. There were movements of Germanic peoples, notably the Franks, into France around the time of the Saxon migration into England. The Germanic ancestry these migrations brought to what is now France would have been Frankish rather than Saxon, and it would have been diluted through mixing with the local populations. It thus seems implausible that ancestry in the UK arising from the Saxon migrations would be better captured by FRA17 than by people now living in the homeland of the Saxons (represented by GER3) – the contribution of FRA17 is about threefold that of GER3. Finally the geographic pattern of FRA17 contributions differs from that of GER3 (which we see as definitely Saxon), in being relatively much higher in the Scottish and Orkney clusters. This is difficult to reconcile with them arriving as part of the same migration event, and the substantial contribution of FRA17 in Scotland and Orkney, relative to GER3, is more likely to reflect an earlier influx into the UK, and increased time to spread geographically. There are similar, though even stronger, arguments against the FRA17 contributions resulting from either Norse or Danish Viking settlement.

We thus conclude that the substantial FRA17 contribution to the UK clusters reflects migration events after those of GER6, BEL11, and FRA14, but largely before the Roman occupation of Britain. It might well represent a steady influx of migrants over long periods before, and even during, the Roman occupation from those areas in France close to the UK coast. Other possibilities would be migration and then growth within the UK associated with particular technologies, including agriculture, but in this case a separate explanation is needed for the lack of contribution of this group in Wales.

Spain

Some earlier analyses of genetic evidence from single marker systems have argued for a Spanish source for ancient British populations, particularly in the west[17]. We see contributions to the ancestry profiles of all the UK clusters from group SFS31 which is sampled in central France and in Spain (principally Barcelona). These contributions range from a low of 1.2% in the large cluster in central and southern England (red squares), to the three highest values ranging from 5.3% to 7.1% for the three Welsh clusters. Whilst caution is needed in interpreting the low levels of contribution from SFS31, this pattern is consistent with limited early migrations, from these areas of Europe, preferentially to the western coastal regions of the UK.

Our data has limitations, in that our sampling in Spain is limited geographically, and includes very few samples from the most natural geographical source regions for Britain, namely Galicia, northern Spain, or the Basque country. If these regions did contribute substantially to British ancestry, we would expect that our approach for estimating ancestry profiles would choose the best surrogates for them in our data, which is likely to be the geographically closest of the groups in our analyses, namely SFS31. Analyses could be further complicated by possible admixture of North-African migrants with Spanish populations subsequent to any movements into the UK. Thus, while our data supports some low level of ancestry from southern France/Spain in ancient British populations it is hard to reconcile with major contributions to modern British ancestry from these regions. More extensive sampling from modern Spain could further clarify this issue.

Italy

We see no contribution to the ancestry of UK clusters from groups in modern Italy. This is not surprising. As noted earlier, there was limited influx of people into Britain during the Roman conquest, and a large existing population. Those who did arrive were mainly Roman soldiers from regions of modern-day France, Germany, and the low-countries. Very few soldiers in the Roman army in Britain were from Rome or modern-day Italy[1].

Ancient Population Structure

We have identified three European groups likely to represent the earliest surviving substantial migrations into Britain, namely GER6, BEL11, and FRA14. Several other groups may also have contributed ancestry around similar times, although at lower levels (see discussion above for caveats): possibly DEN18 (Denmark), SFS31 (southern France/Spain), collectively several of the groups in Norway; and also two Swedish groups. Focussing, however, only on the three major contributing groups, GER6, BEL11, and FRA14, allows us to assess UK population structure after these early migrations but before subsequent migrations. Direct comparison of the contributions from these three groups is complicated by that fact that later migrations may dilute them more in some parts of the UK than in others. For example, the Saxon migrations in the second half of the first millennium introduced DNA from additional source groups preferentially into what is now England and not into Wales. Even if there had been similar levels of ancestry from one of the earlier groups in clusters in Wales and in England, before the Saxon migrations, these levels would be necessarily lower in the English clusters after the Saxon migrations.

In order to understand better the relative contributions to the early population of the UK from the three groups GER6, BEL11, and FRA14, we have undertaken a separate analysis which removes the ancestry contributions from all other groups

and renormalizes the contributions from GER6, BEL11, and FRA14 so they sum to unity. Under the assumption that these three groups represent the earliest migrations, these renormalized contributions estimate the relative contributions in each of the modern day UK clusters from these three "early-migrant" groups. The results are displayed in Extended Data Fig. 6b. The group FRA14 has its highest contributions in all western clusters (*Cornwall*, the three Welsh clusters, the cluster spanning Northern Ireland and western Scotland), while the other two groups have highest contributions in the groups in England (except *Cornwall*), with the ancestry contribution decreasing as one moves away from the clusters in south east and central England. This is consistent with the suggestion from archaeological evidence (see above) of two routes of settlement into the UK after the last glaciation[1], one (best represented in our data by FRA14) by sea up the Atlantic coast of Europe into western Britain (Cornwall and Wales) and Ireland, and the other (best represented in our data by GER6 and BEL11) by land or sea routes into England from the south-east. Under this scenario, migration within Britain since these early migrations then spread DNA from each contributing group throughout the UK, without completely ameliorating the signal of the initial migrations into different areas. Inclusion in this renormalization analysis of the other groups with low-level contributions to all the UK clusters supports the patterns for the three major contributing groups (data not shown).

The preceding analyses suggest that the British population has exhibited population structure since after the migration events that introduced the first sets of ancestors of the modern population. It thus seems problematic to speak of a single "Ancient British" population. Because they have had least dilution from more recent migration events, the samples in our study from Wales carry the highest proportion of ancestry from the early migrations.

'Little England Beyond Wales' or 'English Pembrokeshire'

Our analyses within the UK identified two distinct clusters in south Wales around the county of Pembrokeshire. While they overlap geographically, Fig. 1 shows that one tends generally to correspond to more northerly locations (*N Pembrokeshire*) than the other (*S Pembrokeshire*). The somewhat larger contribution (Fig. 2) to the more southerly *S Pembrokeshire* cluster from BEL11, located in modern Flanders, is consistent with the known Flemish and English settlement of this area in the 12th century. A linguistic barrier (the so-called Landsker line) in Pembrokeshire until relatively recently[18], with English spoken to the south, and Welsh to the north, is likely to have fostered genetic isolation of these two groups. The region to the south of the Landsker line is colloquially referred to as 'Little England Beyond Wales' or in Welsh as 'English Pembrokeshire'. There is also a larger contribution from DEN18 (Denmark) to the *S Pembrokeshire* cluster, consistent with observations of Danish place names in south Wales[19].

Assessing the Accuracy of the Ancestry Profiles

We undertook a number of simulation studies, generating data with similar properties to the actual data, to assess the accuracy of the estimated ancestry profiles (see Methods for details). These suggested good accuracy of the major components of our estimated ancestry profiles. In particular we simulated individuals for three different admixture scenarios: (1) Italy and Northern Germany; (2) North Wales and Norway; and (3) North Wales and Denmark. The first scenario is a test of our model's ability to infer proportions and sources of

admixture when mixing distinct European groups sampled in our data, including the group (GER3) that our real data analysis and interpretation suggests may be representing past Anglo-Saxon migrants. Simulations (2) and (3) take samples from the *N Wales* cluster, which we infer has little evidence of DNA influx related to the Norwegian Vikings and Anglo-Saxons, and mix them with groups containing primarily individuals sampled from Norway (simulation 2) or from Denmark (simulation 3). These simulations mimic admixture between an earlier UK group and Norwegian Viking or Anglo-Saxon settlers, respectively. Simulation (2) further assesses our model's ability to distinguish two distinct Norwegian sources of admixture from among 12 different groups primarily containing samples from Norway.

For each of these scenarios we test a further three possibilities for the proportion of the admixing groups: 10, 25 and 50 per cent for Northern Germany, Norway and Denmark in scenarios (1), (2) and (3) respectively.

We further performed each of these simulations in two ways: one based on the real data chromosomes and the other on a forwards-in-time simulation model. See Methods for details.

The full results are given in Supplementary Table 6. For reasons discussed in the Methods section, the performance of our approach in these simulations is likely to be an under-representation of the performance of our approach in the real data analysis. Given the subtlety of the genetic differences our model is trying to distinguish in this study, it is possible the performance loss will be significant. Furthermore, we only used a relatively small number of individuals from each of these Norway, Denmark, and northern German groups to simulate, because we wanted to ensure a sufficient number of remaining individuals from each to use for inferring the mixing group. As a consequence, the number of simulated individuals we generated is rather small, consisting of only 25 or 40 individuals per simulation, compared to our real data analysis where, for example, the *Cent./S England* cluster has 1,044 individuals. We expect the increased sample size in our real data to improve our inference relative to these simulations; substantially so in some cases such as *Cent./S England*.

Despite the caveats regarding performance given above, our simulation results are encouraging for demonstrating our approach's ability to infer the proportions of DNA attributable to different European groups, even in our limited simulation setting. Considering first the simulations based on the real data we note the following. For the simulations in (1) consisting of 10%, 25% and 50% admixture from the northern German group GER3 and the remainder from the Italian group ITA36, our model infers a contribution of 6.4%, 20.0% and 36.9%, respectively, from (the remaining individuals not used to simulate in) GER3 and 86.2%, .71.6% and .45.1%, respectively, from (the remaining individuals not used to simulate in) ITA36. This demonstrates our model's ability to identify and reliably quantify distinct sources of admixture among our sampled European groups, even with only 25 admixed individuals. For the simulations in (2) consisting of 10%, 25% and 50% admixture from the Norwegian groups NOR72/NOR71, our model infers a total contribution of 14.1%, 25.7% and 44.7%, respectively, when summing the contributions from groups NOR53, NOR61, NOR63, NOR64, NOR71, NOR72, NOR80, NOR81 NOR85, NOR90, NOR102 AND NOR139; all groups containing samples predominantly from Norway. The inferred contributions are 9%, 19.7% and 37.8%,

respectively, if you consider only the contributions from the groups NOR72 and NOR71 used to simulate, suggesting that our model can accurately identify the precise Norwegian groups involved in admixture events. Finally for the simulations in (3) consisting of 10%, 25% and 50% admixture from the Danish group DEN18, our model infers a contribution of 13.3%, 20.4% and 39.5%, respectively, from DEN18, suggesting we are able to accurately distinguish between varying levels of admixture from Denmark (though with perhaps a slight underestimate for higher fractions, when inferring with only 25 simulated individuals). Reassuringly, for each of (2) and (3), the remaining contributions closely mirror our model's inferred contributions from Europe for the *N Wales* cluster.

The results for the forwards-in-time simulation procedure closely matched those discussed above for all nine of the scenarios. Collectively these results lead us to conclude that the ancestry profile analyses are robust.

Differences Between Ancestry Profiles

It is possible for distinct fineSTRUCTURE clusters to have very similar ancestry profiles (e.g. Cumbria and Northumbria, Fig. 2). Two sets of individuals could receive similar contributions from a set of European groups (leading to similar ancestry profiles) but then evolve separately (leading to different patterns of ancestry, and thus to distinct clusters in fineSTRUCTURE). One can calculate the total variation difference between the ancestry profiles of a pair of clusters (TVD_{AP}; see Methods). TVD_{AP} can be interpreted as a measure of the strength of the differences in ancestry of the two clusters.

Supplementary Table 5 gives TVD_{AP} for all pairs of ancestry profiles for the 17 UK clusters used in our main analyses, and gives a p-value (based on a permutation test) for the significance of the differences observed (see Methods for details). In spite of the visual similarity of many of the ancestry profiles, most of the pairwise comparisons of larger clusters show significant differences. The exceptions tend to be for clusters in similar geographical regions. The power to detect significant differences in comparisons of smaller clusters is more limited, making the non-significant p-values harder to interpret, but we note that many of these are in relatively close geographical proximity, making similar contributions to ancestry, and hence similar profiles, more plausible.

Characterising a simulated "Italy and Northern Germany" admixture event using GLOBETROTTER

To test the applicability of GLOBETROTTER in the particularly challenging setting of admixture within Europe, between extremely similar sources, we applied the algorithm implemented in GLOBETROTTER to infer the nature of the 'Italy and Northern Germany' simulation with $N = 25$, $\lambda = 40$, $\beta = 0.25$ (Extended Data Fig. 8). Strong evidence of admixture was seen ($p<0.01$). Although uncertain given the small sample size, the estimated admixture date of 40 generations ago (95% CI of 18-55 generations) was identical to the truth. Notably, the inferred admixture fraction of 24% and the inferred sources (most similar to GER3 and ITA36, exactly matching the admixing sources used for the simulation, and with contributions of 24% and 76% from these sources respectively) were extremely close to the truth. Also notably, the value of δ_{MN} (see Methods) for the best-fitting curve of ~0.0001 implies only very weak information about underlying ancestry is available to GLOBETROTTER for this simulated admixture event at any single locus in the genome, necessitating a genome-scale analysis of multiple individuals to

understand such events. This is comparable to strength of the signal seen in the UK analyses discussed below.

We repeated this for the two other 'Italy and Northern Germany' simulations with $N = 25$, $\lambda = 40$, $\beta = 0.10$, and $N = 25$, $\lambda = 40$, $\beta = 0.50$. As before, in each case strong evidence of admixture was observed ($p<0.01$). For the case when $\beta = 0.10$ the inferred admixture fraction was 14%, and the inferred sources were again most similar to GER3 and ITA36, with contributions of 14% and 86% from these sources respectively. The estimated admixture date was 46 generations ago (95% CI of 29-66 generations). For $\beta = 0.50$ the inferred admixture fraction was 47% (the inferred sources were most similar to GER3 and ITA36, with contributions of 47% and 53% respectively) and the estimated admixture date was 54 generations ago (95% CI of 38-70 generations). In both cases the confidence intervals are large (due to the small sample size N = 25), but overlap the truth.

Dating Admixture Events in Orkney and South East England

As described in the main text, we applied the algorithm implemented in GLOBETROTTER to infer the nature of any possible admixture events that may have contributed to the ancestry profiles we observe for the *Cent./S England* cluster and the three clusters in Orkney (*Westray*, *Orkney 1* and *Orkney 2*). In particular we sought to determine if there was evidence that an admixture event had occurred, and if so, when and in what proportions.

Although our simulations of 'Italy and Northern Germany' (see above) resulted in highly accurate results using GLOBETROTTER, we caution that the extreme subtlety of admixture signals expected in the UK may lead to an identifiability issue[20], predicted from theory, where the admixture proportion cannot be definitively inferred from the data using GLOBETROTTER. In admixture events between different groups, the mixture fraction is identifiable, provided that each source group has at least one admixing population in the appropriate "mixture" decomposition contributing only to that group, and not to the other admixing population. However, in the setting where both admixing groups copy very similar amounts from all sampled populations - a likely issue for our setting of admixture between north west European populations - the groups contributing to each admixing population in the mixture representation might be almost the same, and how they actually divide up cannot be fully identified, equivalent to uncertainty in the admixture fraction. We conservatively assumed this problem would occur in our UK GLOBETROTTER analyses. In this setting, the admixture date, and properties of the differences between the source groups can still be inferred, as can the overall population make-up in terms of a mixture, but precisely how this mixture is divided up between the two populations cannot be fully determined. Thus, we restrict ourselves to discussing properties of the differences between the true admixing sources, the overall makeup of the resulting population, and the admixture date inferred in the UK GLOBETROTTER analyses.

We found strong evidence ($p<0.01$) of admixture in all four of the UK clusters (*Cent./S England*, *Westray*, *Orkney 1* and *Orkney 2*) analysed, with none having any strong evidence of multiple dates of admixture ($p>0.05$), consistent with a single "pulse" of admixture, and providing very strong evidence of recent admixture having influenced these parts of the UK (Extended Data Fig. 9). The 95% confidence intervals for when this admixture occurred were 802-914 for *Cent./S England* and 830-1418, 1082-1530, and 438-1278 for the three Orkney clusters (*Orkney 1*,

Westray and *Orkney 2* respectively). The confidence intervals for the three Orkney clusters, in all cases, overlap (and approximately span) the period of Norse occupation in Orkney (from the late 8th Century to the 15th century). For each UK cluster, the inferred genetic make-up of each source group for the strongest detected event gave results that were largely consistent with the ancestry profiles inferred as described in "Estimating Ancestry Profiles". Specifically, the Orkney clusters *Orkney 1*, *Westray* and *Orkney 2* were inferred to have 25.2%, 22.5%, and 21.8% of their respective DNA in common with European groups primarily containing individuals sampled from Norway. As discussed above, caution must be exercised when considering the inferred admixture fractions, but there is value in considering the differences in the sources that we observe. With this in mind we note that in each of the three Orkney clusters one of the admixing groups is distinguished by sharing more haplotypes with present-day groups found in Norway (especially groups found on the west coast of Norway), while the other group copies more DNA from a range of other European populations including France. This means that we infer people from Orkney as having genomes formed by admixture between one more Norwegian-like groups, and a more cosmopolitan French-like group, approximately 900 years ago. This strongly accords with what one might expect from the history of Norse settlement in Orkney, confirming the value of our approach which makes inference independently of any prior assumptions about the history and genetic make-up of Orkney.

The *Cent./S England* inferred admixture date is older, at around 1200 years ago. This is moderately, but significantly, more recent than the historically accepted time of approximately 1400 years ago (around 600) for the Anglo-Saxon migration into England. This discrepancy is unlikely to be explained by errors in our human generation time (we used 28 years) because an unlikely generation time of 33 years or higher would be required to account for this difference. Instead, an important point is that the date of admixture cannot be earlier than the arrival of a group, but can be later if mixing did not occur for some period (e.g. if the Anglo-Saxon community remained distinct for some period after arrival), or if mixing took place gradually, and initially at a relatively slow rate. The latter case is often difficult/impossible for GLOBETROTTER to distinguish from a single admixture "pulse"[20] and instead GLOBETROTTER produces a date estimate within the range of the period of mixing. Finally, it is possible that a later Viking influx (in the period 800-950), especially of Danish Vikings from similar geographic locations to the Anglo-Saxons (in particular the Angles and Jutes), is contributing some of the observed signal, and pushing the estimated admixture date somewhat towards the present day. The overall inferred makeup of haplotypes in the *Cent./S England* cluster included a 35% contribution from European group GER3, who are found most substantially in northern Germany. This estimate is slightly higher than seen in our original analysis as described in "Estimating Ancestry Profiles", but confirms that there is a minority contribution of GER3 ancestry to the *Cent./S England* cluster. Moreover, the difference between inferred admixing sources indicated one admixing source copying much more from GER3, somewhat more from the Danish group (DEN18), and slightly more from a range of Norwegian groups. The other admixing source was similar to that seen for Orkney, in that it copies more French DNA. This is more consistent with one identified admixture source corresponding mainly to Anglo-Saxons from today's northern Germany and Denmark – because unlike the Anglo-Saxons, no Vikings originated from northern Germany - but could include a smaller contribution from the (Danish or Norse) Vikings.

Consortium Information

Wellcome Trust Case Control Consortium 2†,
International Multiple Sclerosis Genetics Consortium†.

†For information about participants of this consortium (not all of whom are authors of this paper) see Nature 476, 214–219 (2011).

References

1. Cunliffe, B. *Britain Begins*. (Oxford University Press, 2013).

2. Coles, B. J. Doggerland: a speculative survey. *Proc. Prehist. Soc.* **64,** 45–81 (1998).

3. Brown, A. Dating the onset of cereal cultivation in Britain and Ireland: the evidence from charred cereal grains. *Antiquity* **81,** 1042–1052 (2007).

4. Smith, C. The Population of Late Upper Palaeolithic and Mesolithic Britain. *Proc. Prehist. Soc.* **58,** 37–40 (1992).

5. Pryor, F. *Britain BC: Life in Britain and Ireland Before the Romans*. (Harper Perrenial, 2004).

6. Ward-Perkins, B. Dating the onset of cereal cultivation in Britain and Ireland: the evidence from charred cereal grains. *Engl. Hist. Rev.* **115,** 513–533 (2000).

7. Hinde, A. *England's Population: A History Since the Domesday Survey*. (Hodder Arnold, 2003).

8. The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476,** 214–219 (2011).

9. Clarkson, T. *The Makers of Scotland: Picts, Romans, Gaels and Vikings*. (Birlinn, 2012).

10. Bardon, J. *A History of Ulster*. (Blackstaff Press, 2001).

11. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–668 (2007).

12. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).

13. Wilson, J. F. *et al.* Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5078–5083 (2001).

14. Wells, R. S. *et al.* The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10244–10249 (2001).

15. Capelli, C. *et al.* A Y chromosome census of the British Isles. *Curr. Biol.* **13,** 979–984 (2003).

16. Goodacre, S. *et al.* Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. *Heredity (Edinb).* **95,** 129–135 (2005).

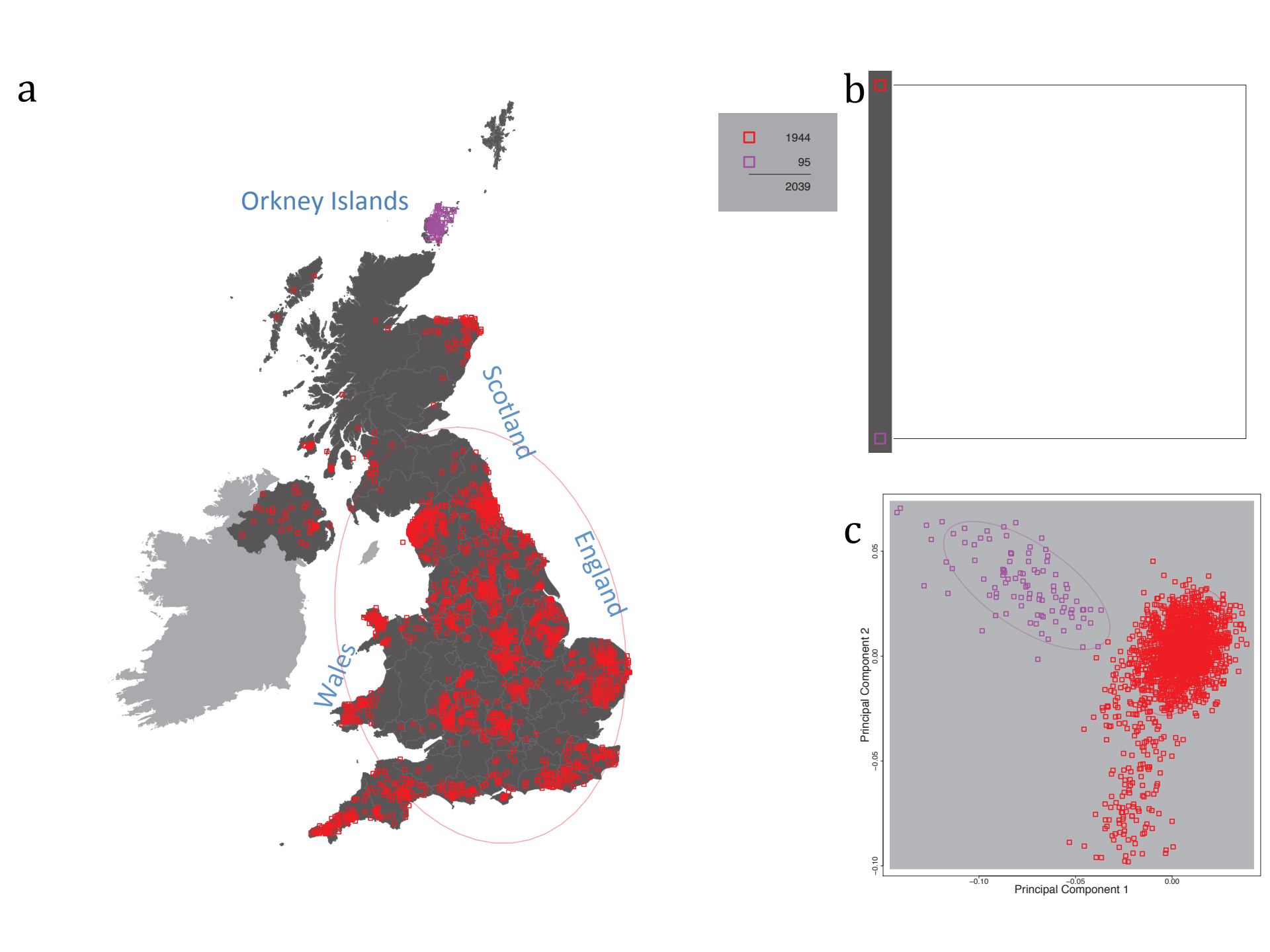
17. Oppenheimer, S. *The Origins of the British: a Genetic Detective Story*. (Constable and Robinson, 2006).

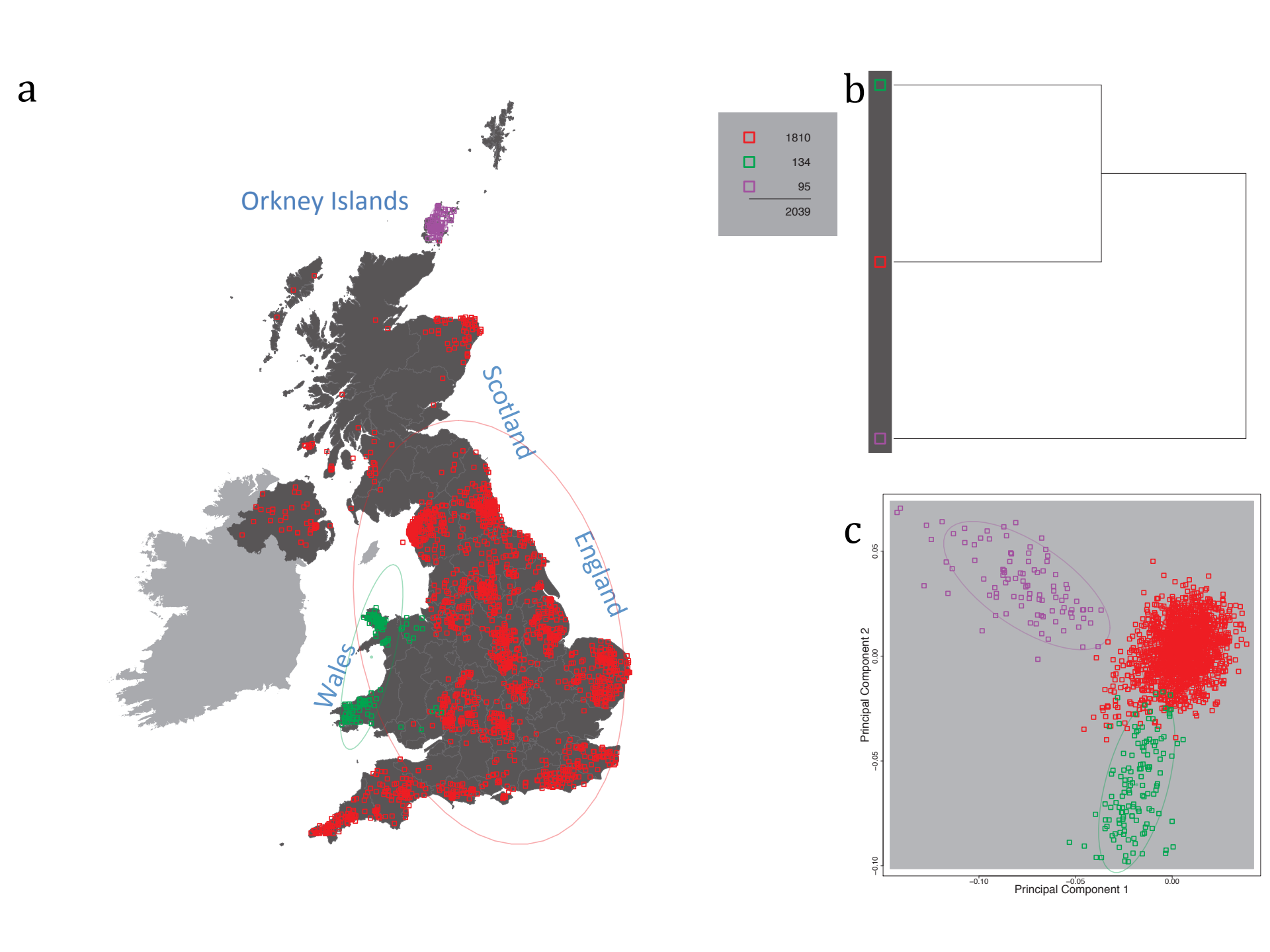
18. Awbery, G. M. The term "landsker" in Pembrokeshire. *J. Pembrokesh. Hist. Soc.* **4,** 32–44 (1991).

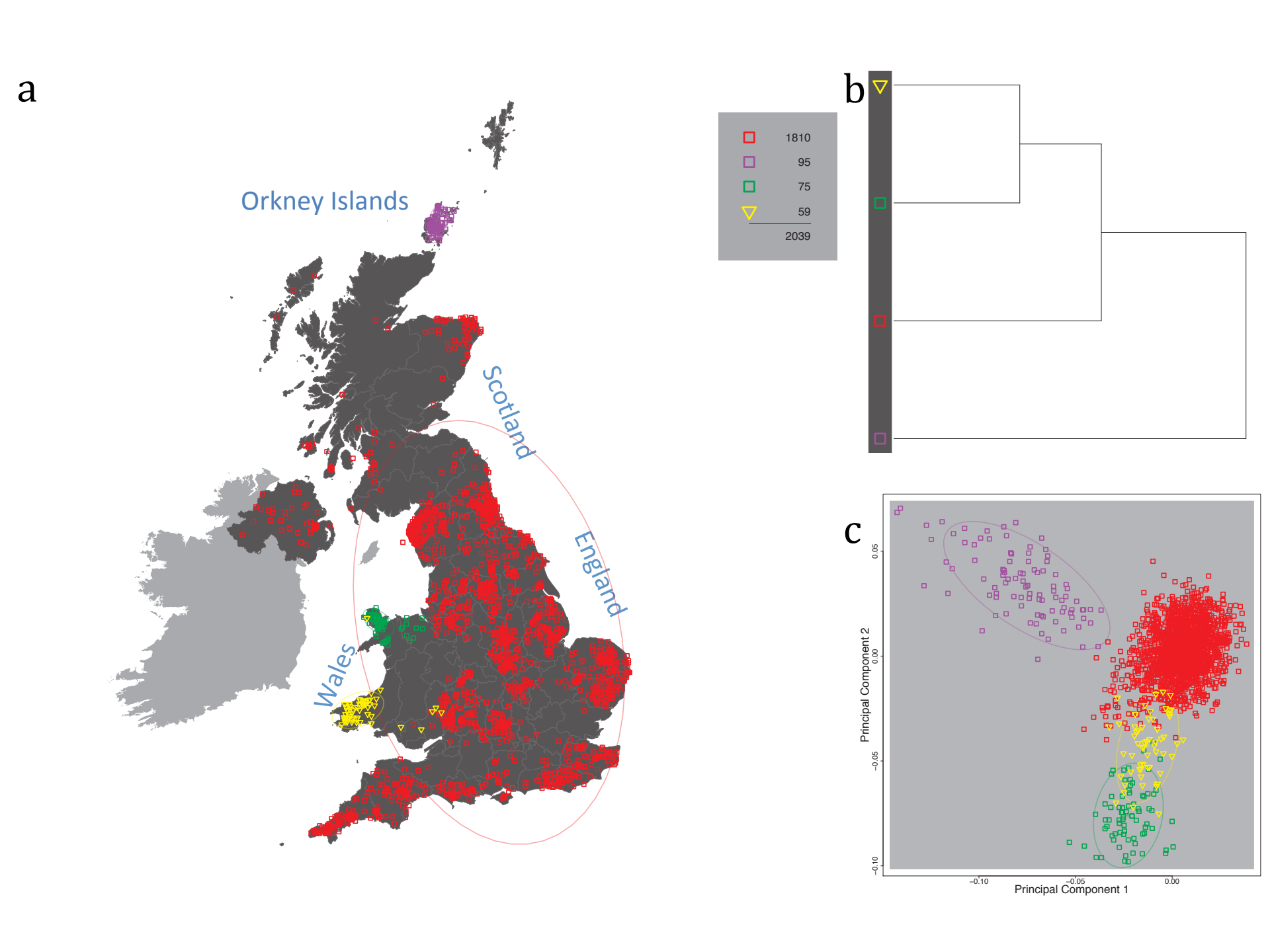
19. Haywood, J. *The Penguin Historical Atlas of the Vikings*. 78–79 (Penguin, 1995).

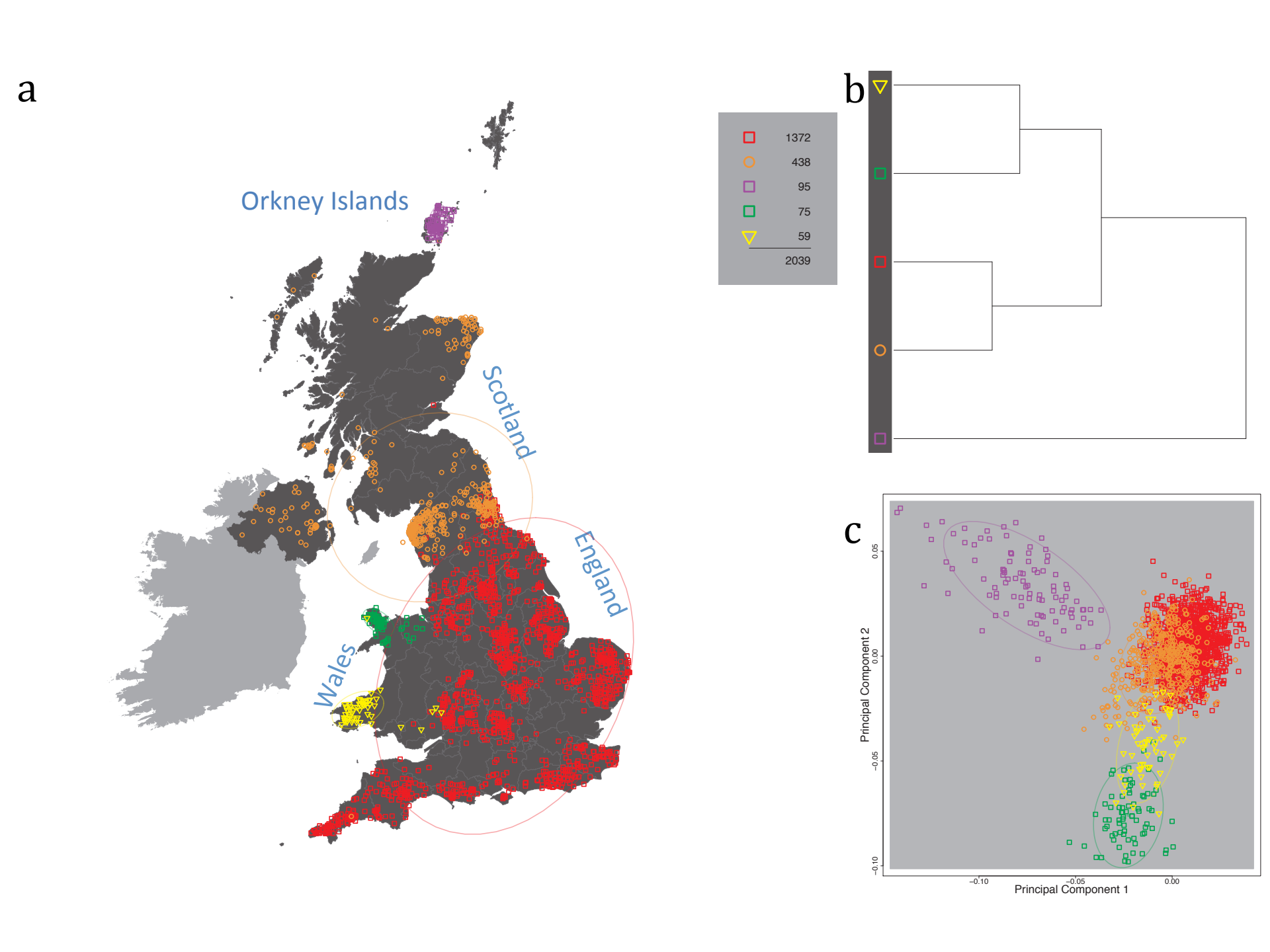
20. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343,** 747–751 (2014).

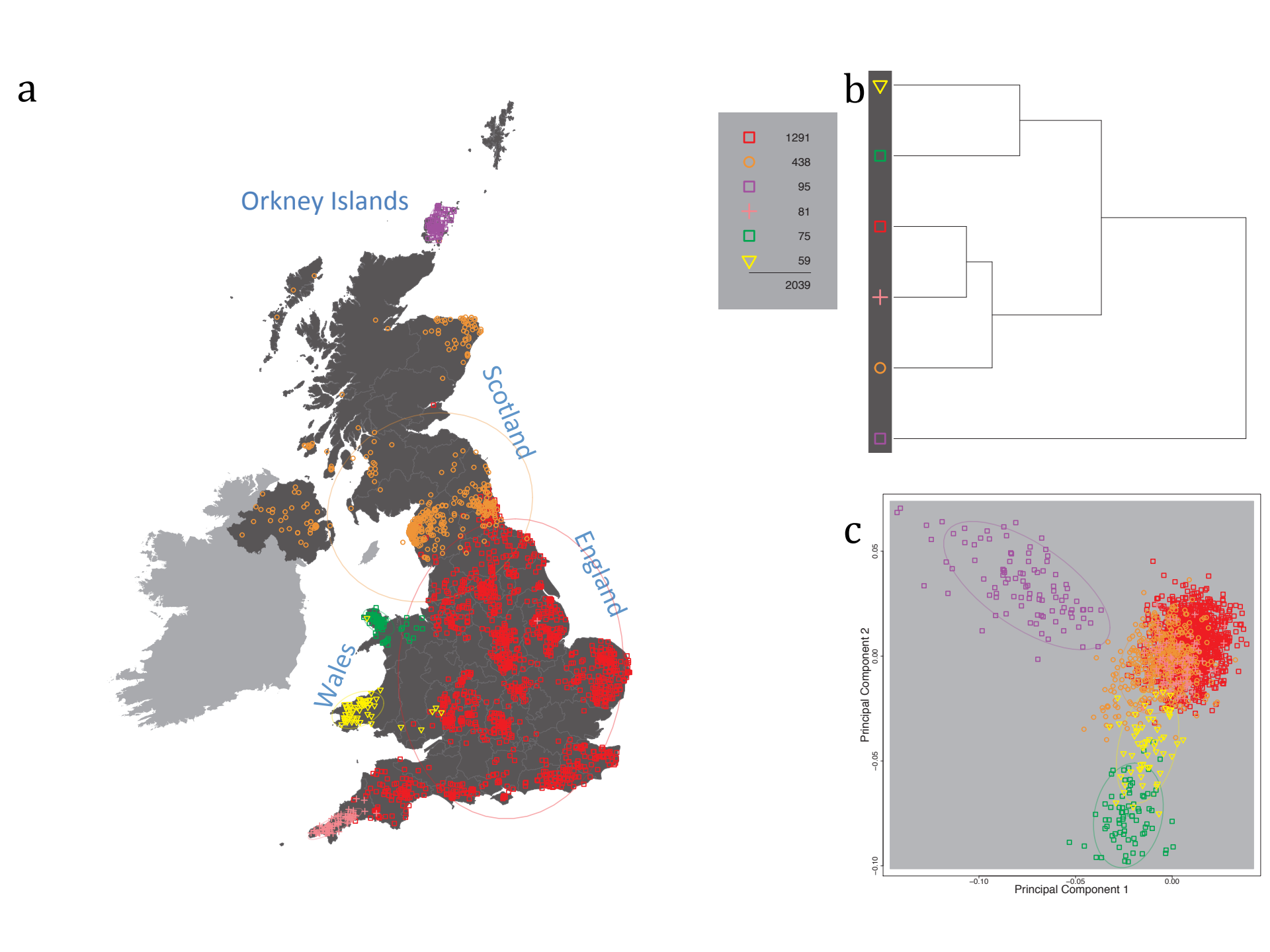
Supplementary Figure 1.1 - 1.24 | Genetic clusters in the UK inferred by the fineSTRUCTURE analysis at all levels of the hierarchical clustering. Each of the plots 1.1 – 1.23 shows exactly the same information, but for different numbers of clusters, from 2 to 24 in order, determined by the hierarchical clustering analysis. At the level of 24 clusters every cluster has at least 10 members. This is not the case for finer levels of clustering and for brevity these levels are omitted. The final figure, 1.24 shows the final clustering by fineSTRUCTURE, with 53 clusters. **a**, The UK map depicts the clustering of the 2,039 UK individuals into clusters on the basis of genetics alone. Each symbol corresponds to one of the sampled individuals and is plotted at the centroid of their grandparents' birthplace. Each genetic cluster is represented by a unique combination of colour and plotting symbol, with individuals depicted with the symbol of the cluster to which they are assigned. The ellipses centred on each cluster give a sense of the extent of the cluster by showing the 90% probability region of the two-dimensional t-distribution (5 degrees of freedom) which best fits the locations of the individuals in the cluster. No relationship between clusters is implied by the colours/symbols. In addition there is a table at each level that displays the number of samples in each of the inferred clusters. **b**, A tree depicting the order of the merging of the clusters in the hierarchical clustering. The cluster symbols are the same as shown in **a**. The lengths of the branches relate to changes in the posterior of the statistical model underlying fineSTRUCTURE. They do not relate directly to time or other measures of genetic distance so caution is needed in their interpretation. Some additional length is added to the tips of the tree for clarity. **c**, The UK samples plotted against the first two principal components as determined in the genome-wide principal components analysis. For comparison, each individual is depicted by the same symbol as in the fineSTRUCTURE analysis depicted in **a**. The ellipses are drawn as in **a**.

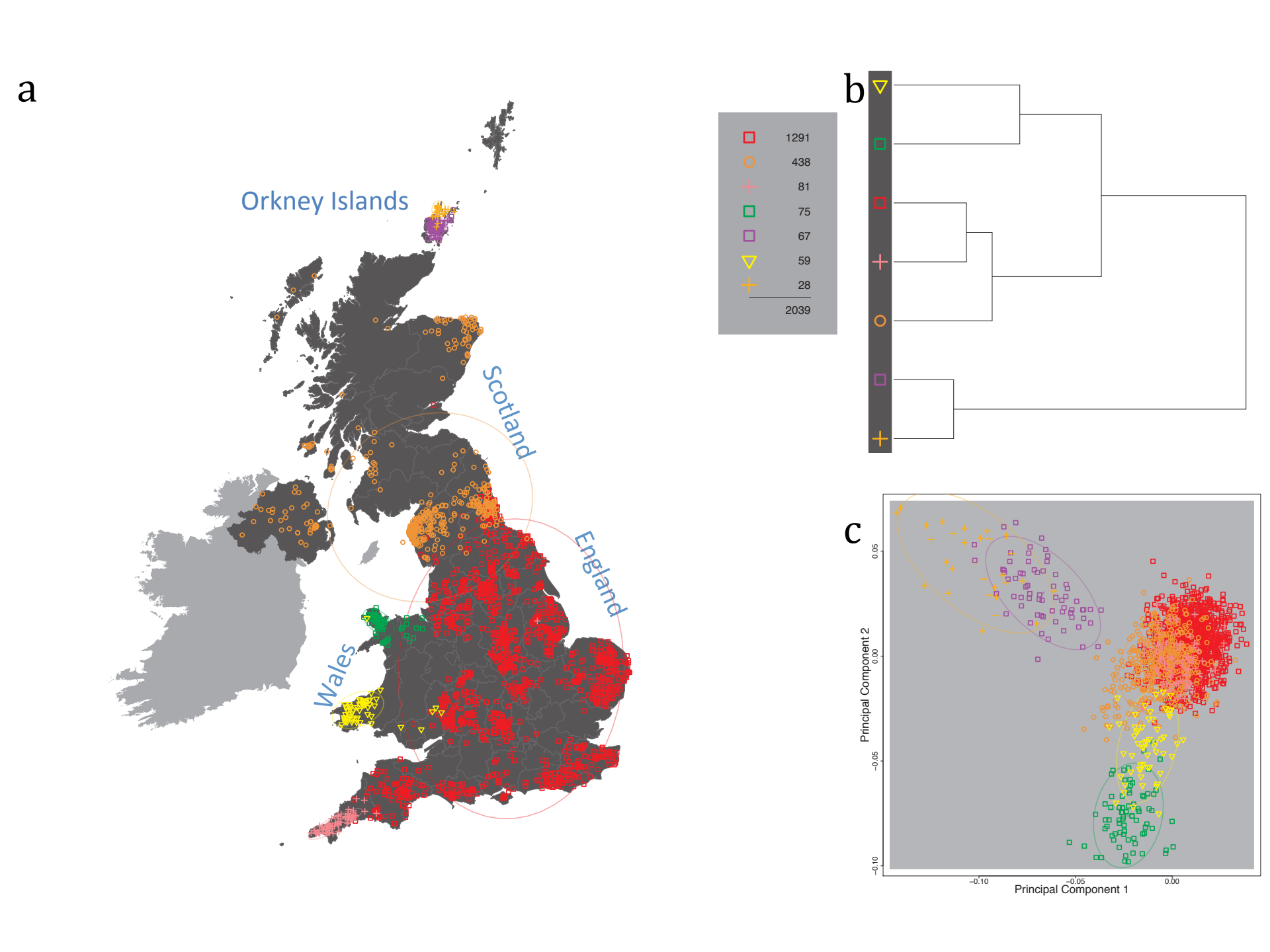

a
Orkney Islands
Scotland
England
Wales
1944
95
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

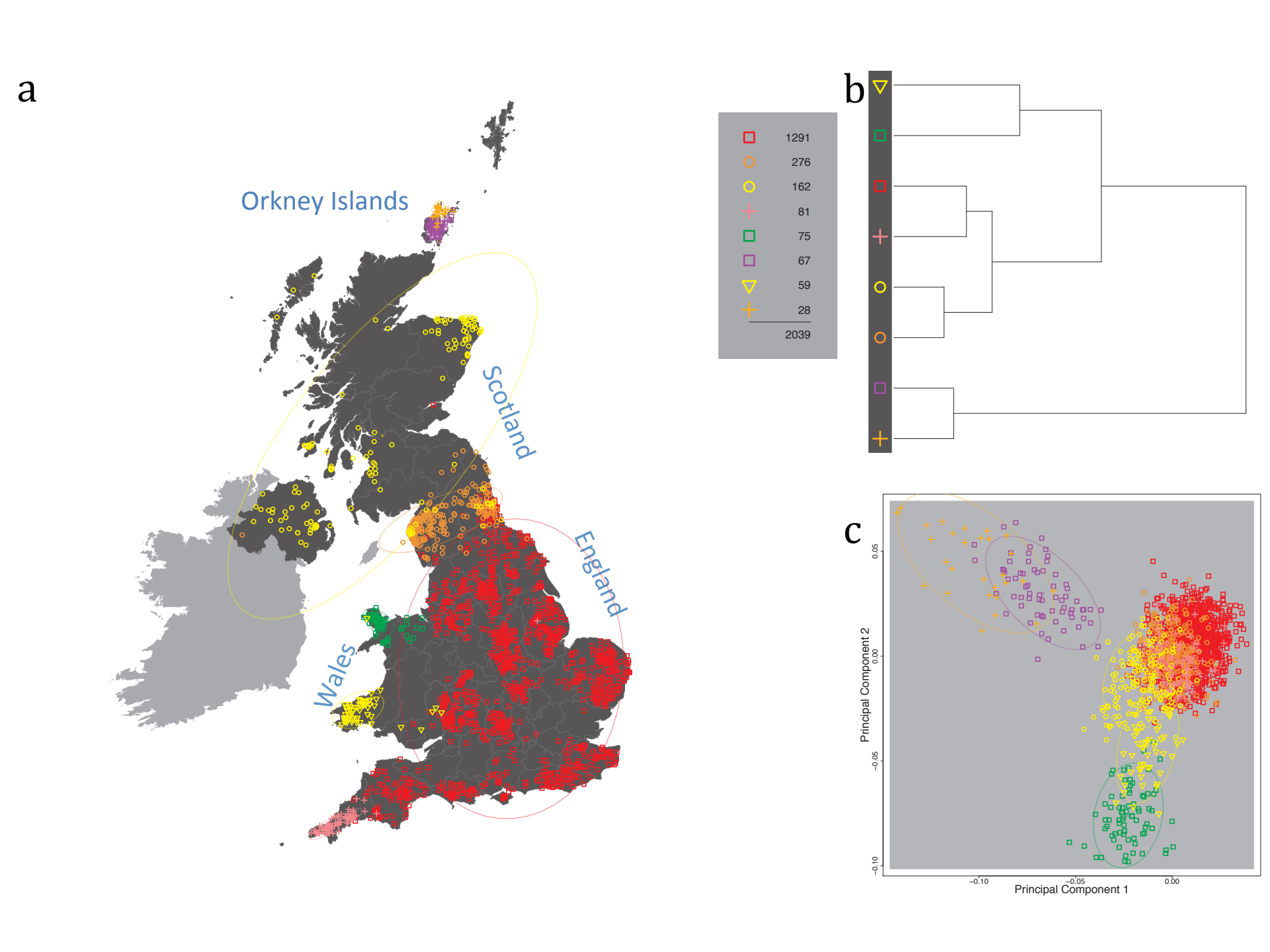

a
Orkney Islands
Scotland
England
Wales
1810
134
95
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

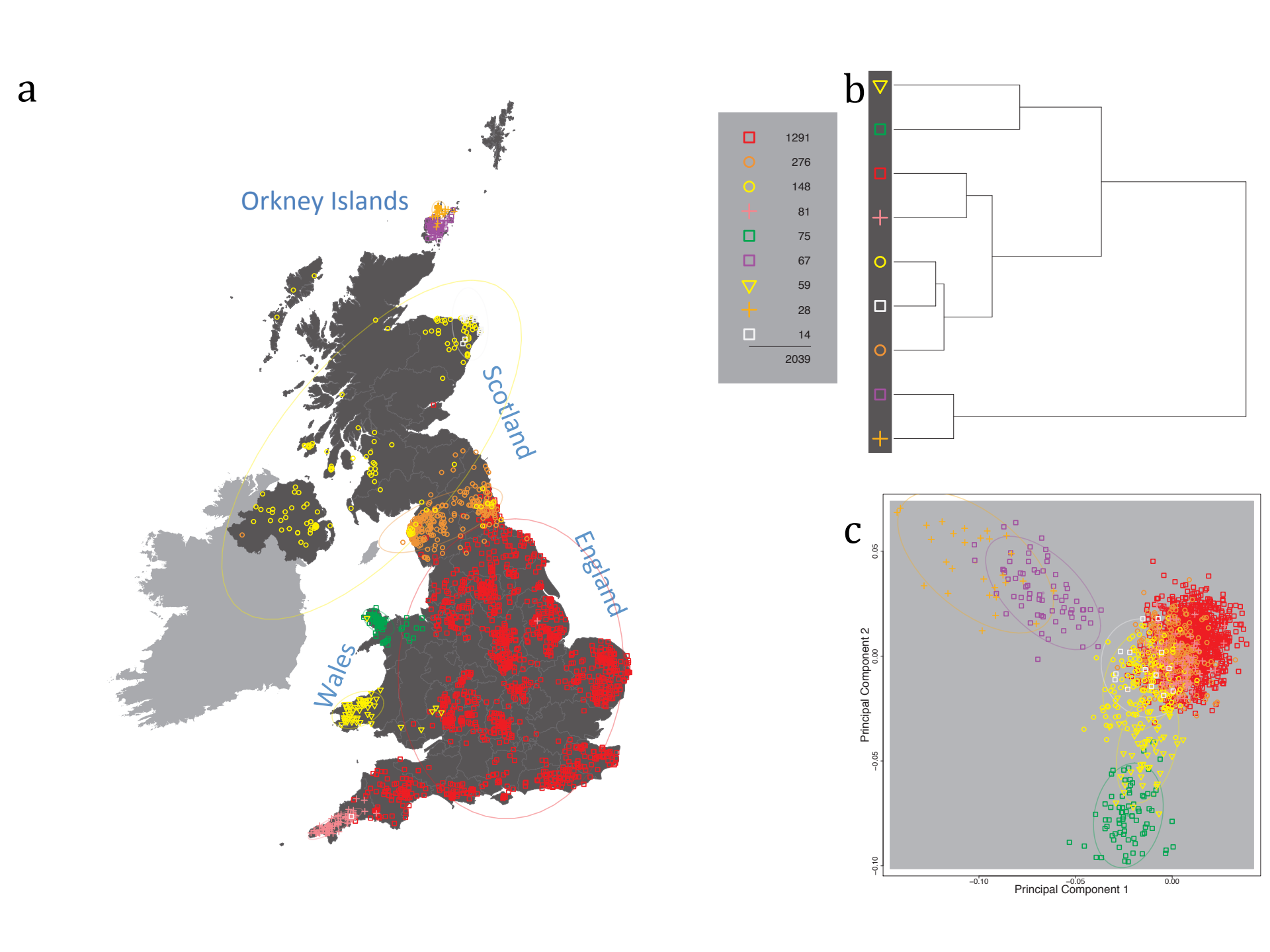

a
Orkney Islands
Scotland
England
Wales
1810
95
75
59
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

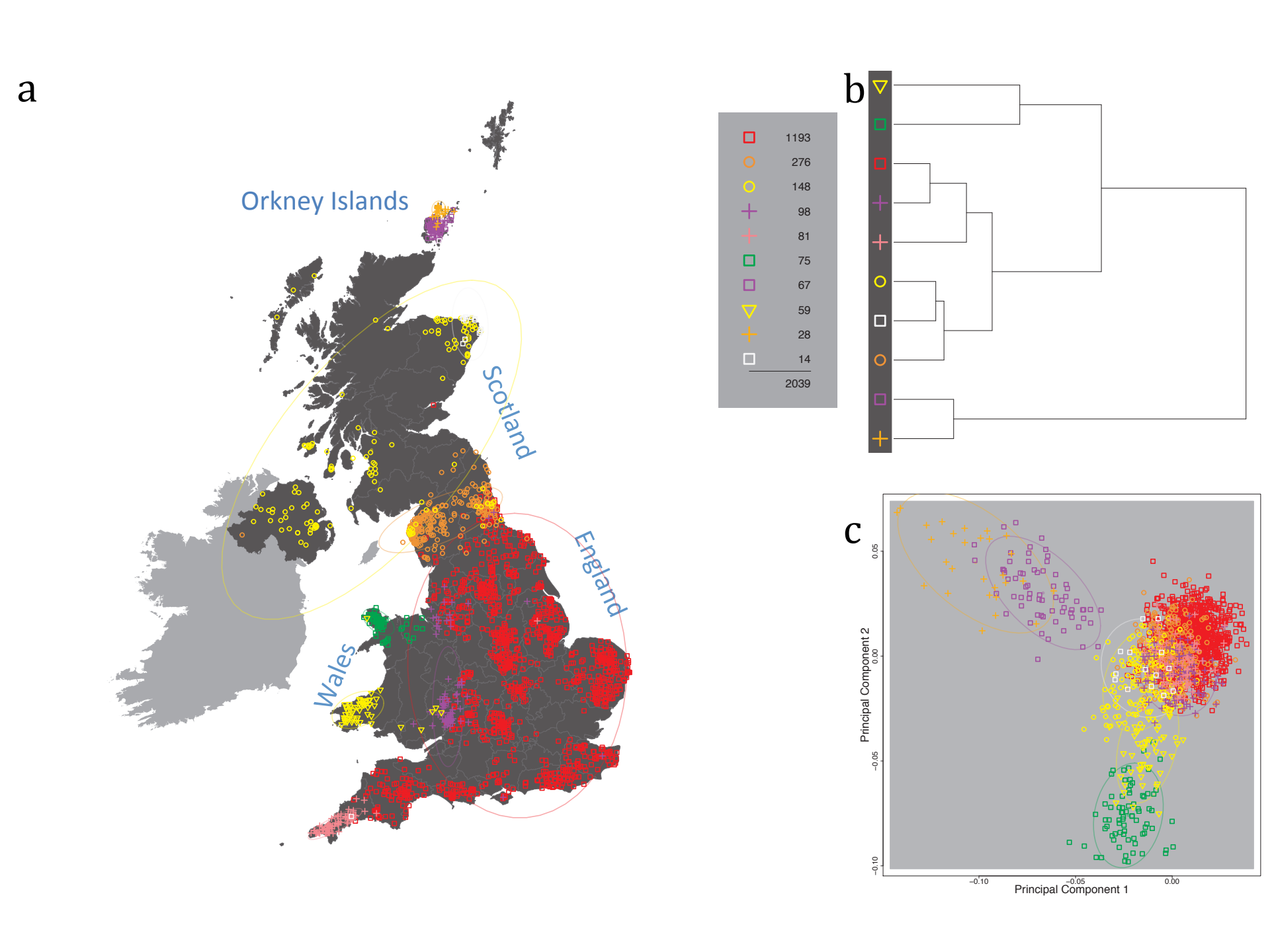

a
Orkney Islands
Scotland
England
Wales
1372
438
95
75
59
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

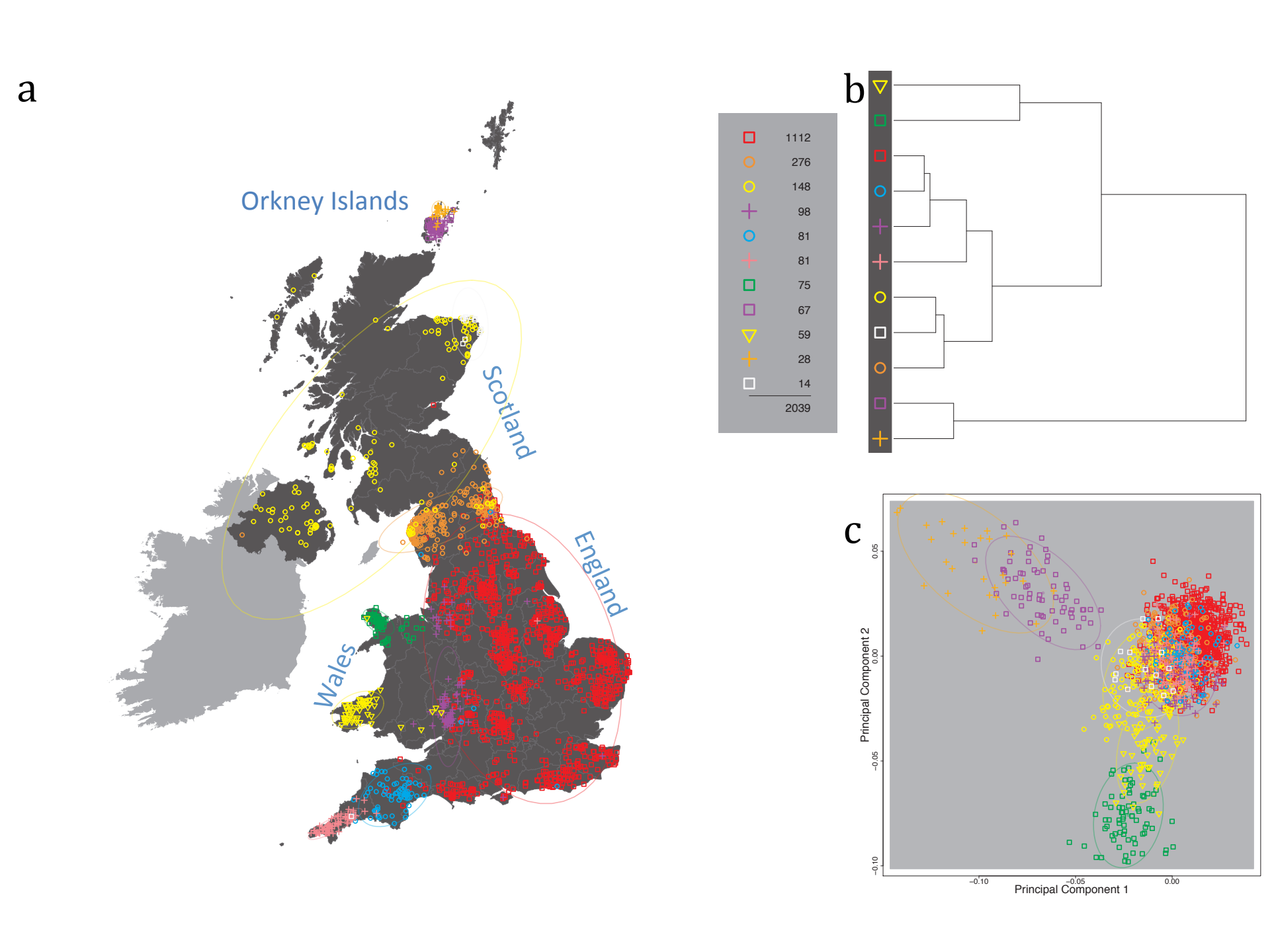

a
Orkney Islands
Scotland
England
Wales
1291
438
95
81
75
59
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

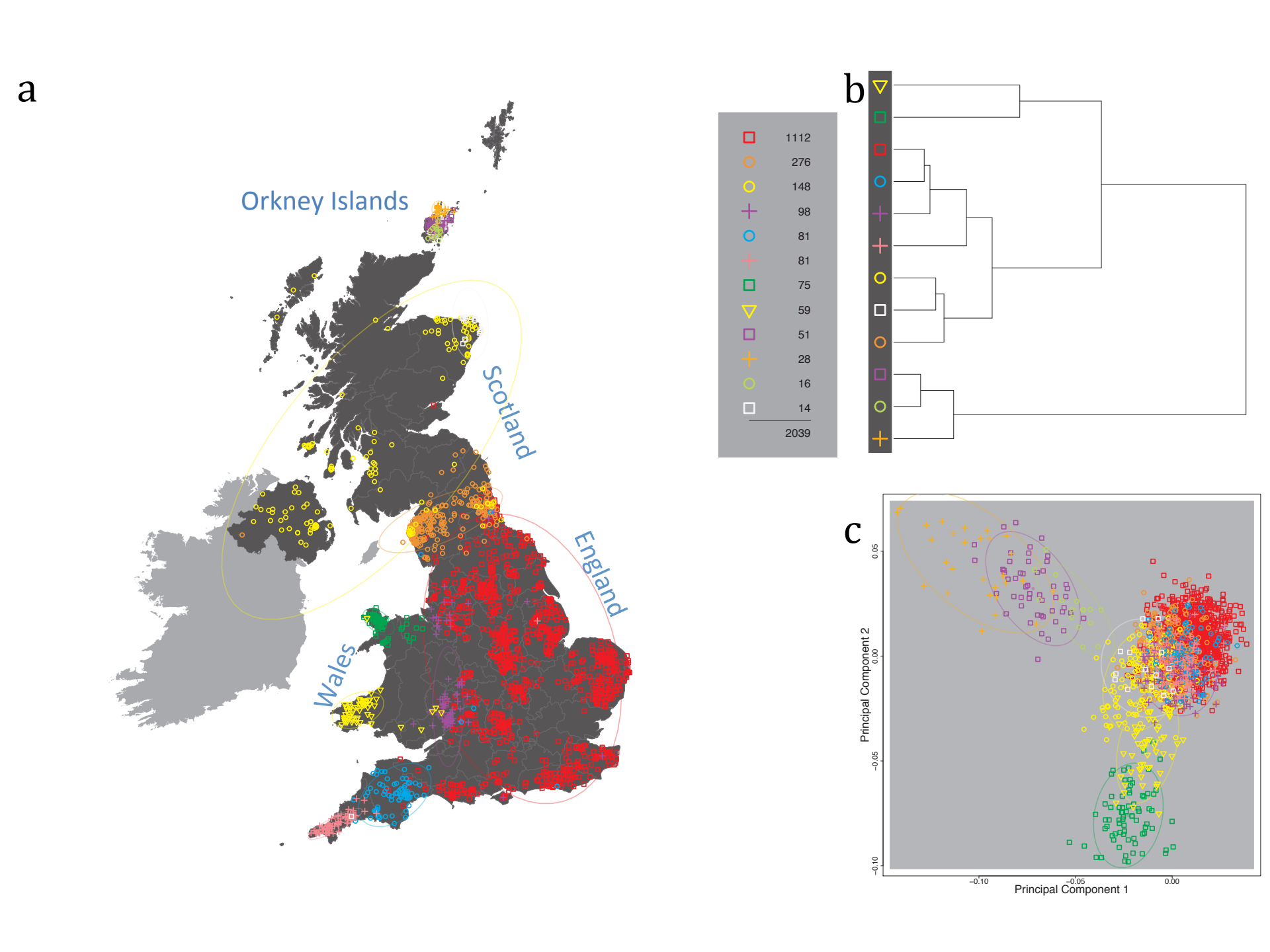

a
Orkney Islands
Scotland
England
Wales
1291
438
81
75
67
59
28
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

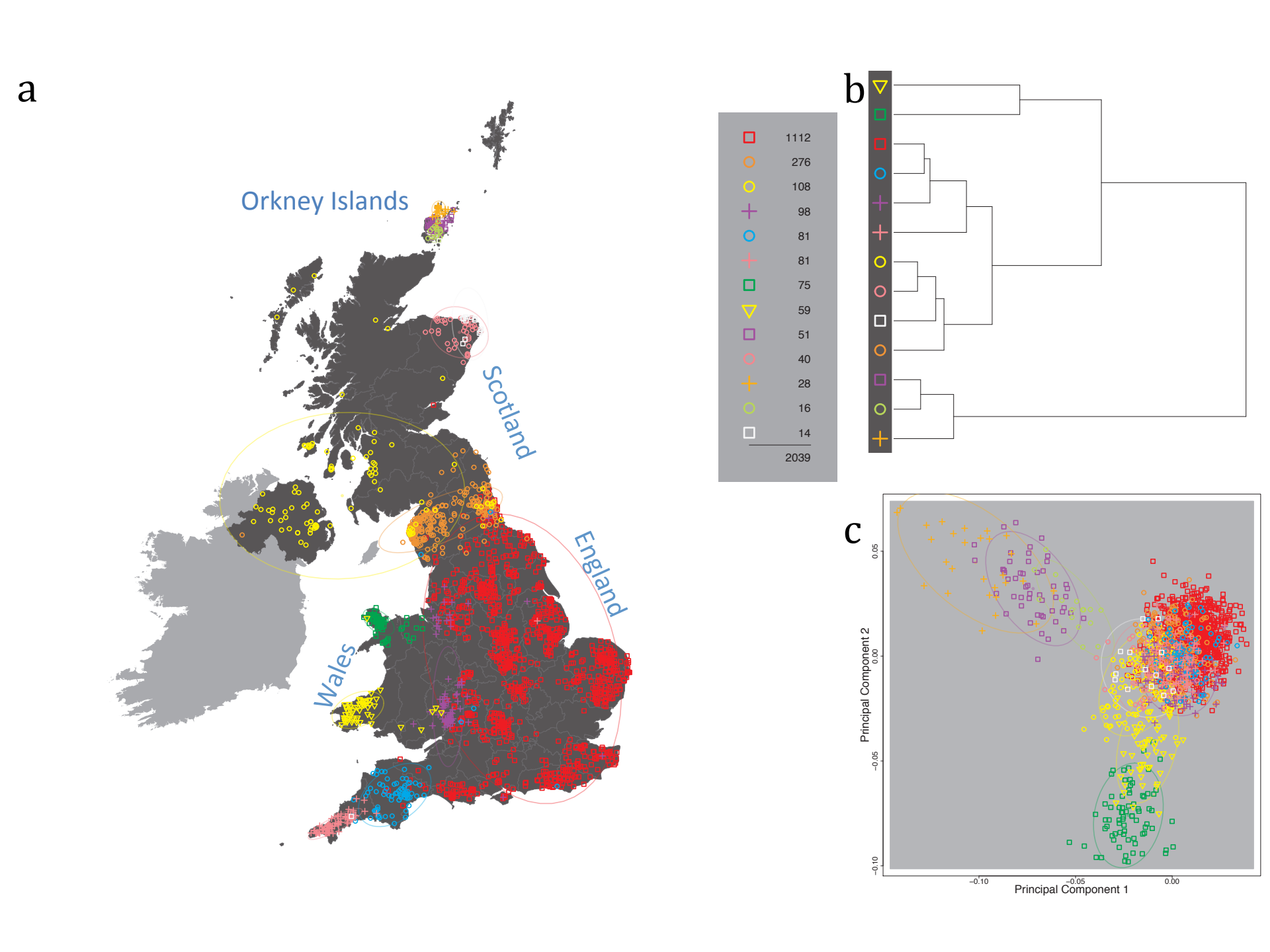

a
Orkney Islands
Scotland
England
Wales
1291
276
162
81
75
67
59
28
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

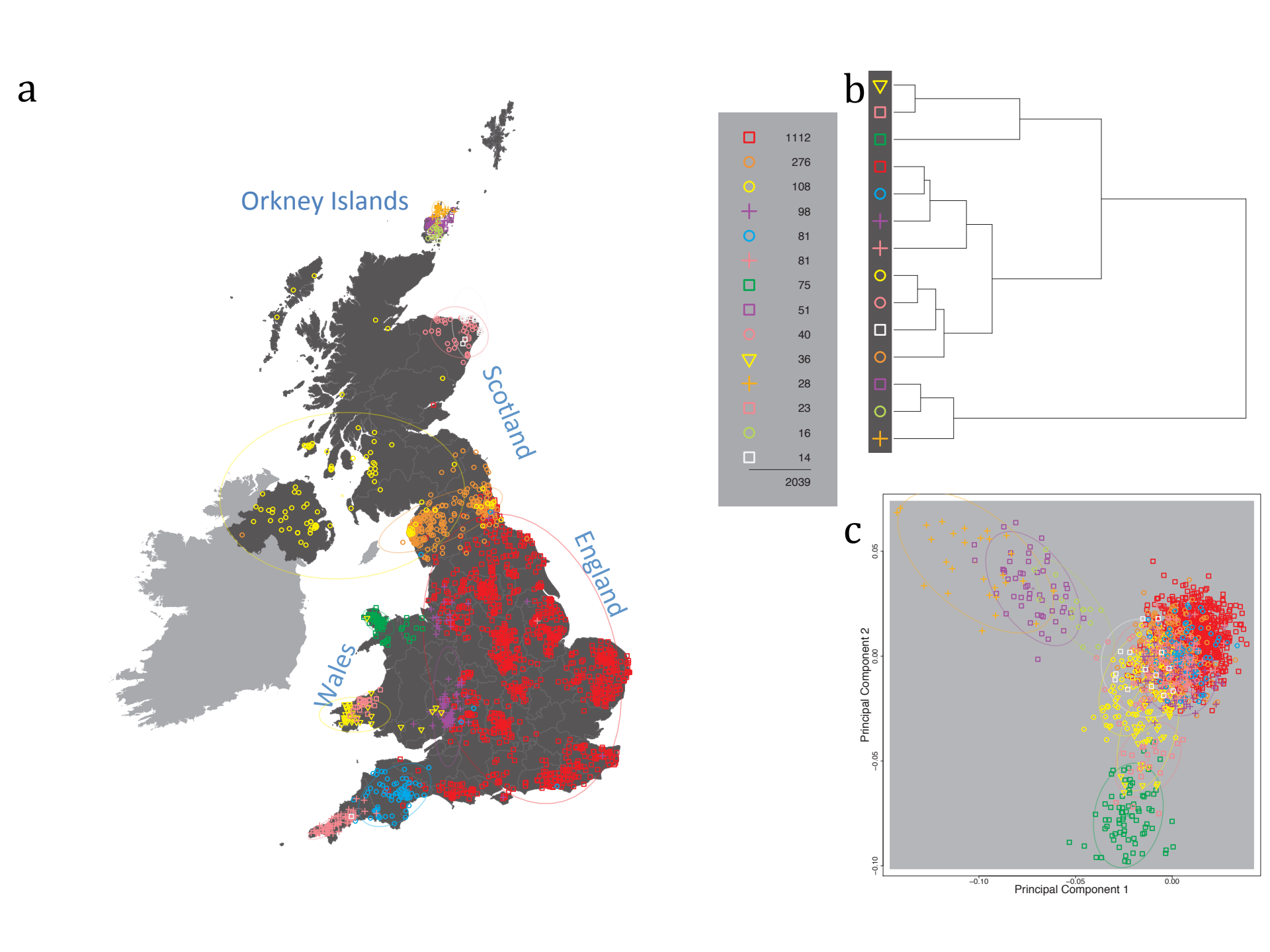

a
Orkney Islands
Scotland
England
Wales
1291
276
148
81
75
67
59
28
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

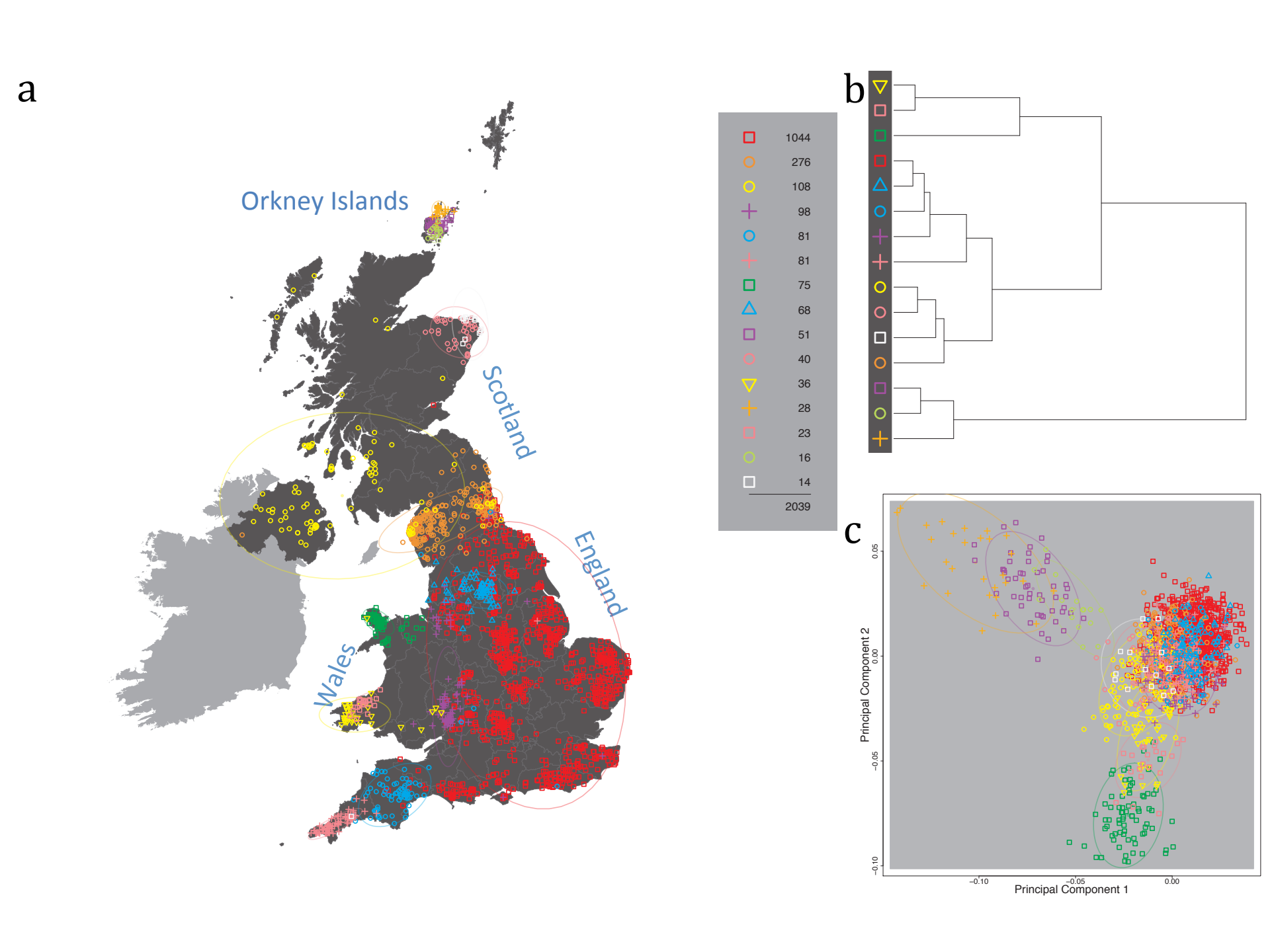

a
Orkney Islands
Scotland
England
Wales
1193
276
148
98
81
75
67
59
28
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

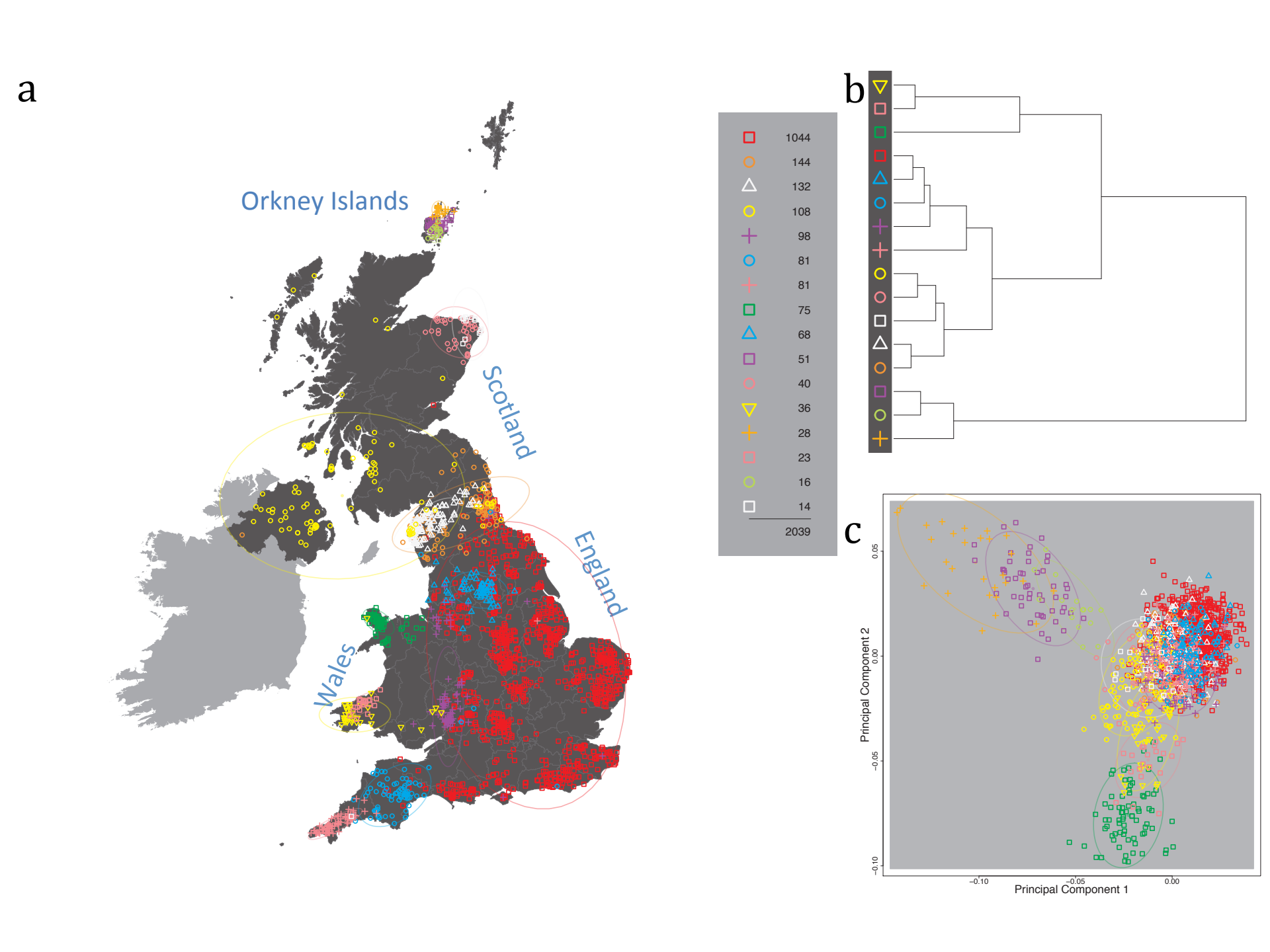

a
Orkney Islands
Scotland
England
Wales
1112
276
148
98
81
81
75
67
59
28
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

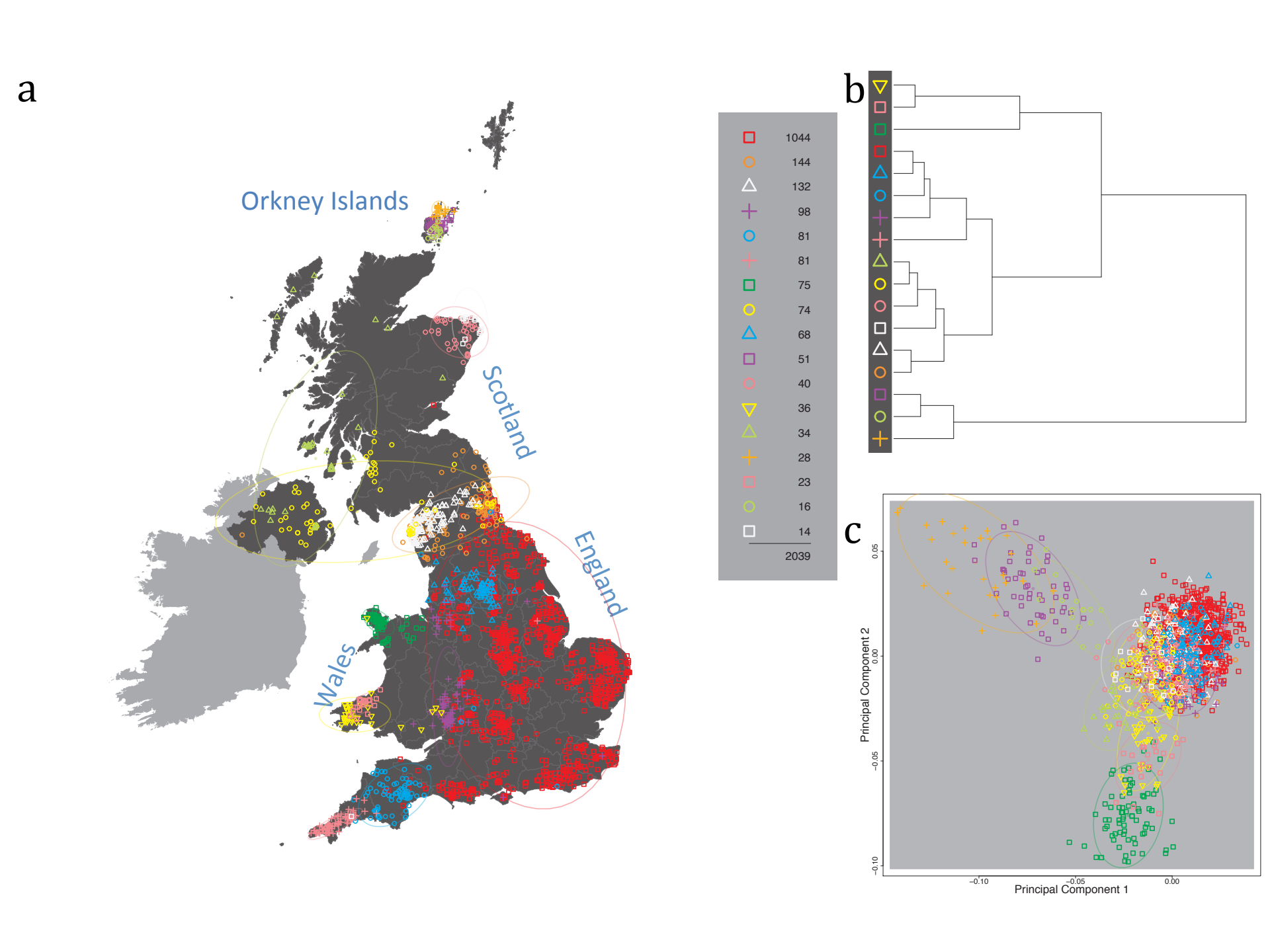

a
Orkney Islands
Scotland
England
Wales
1112
276
148
98
81
81
75
59
51
28
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

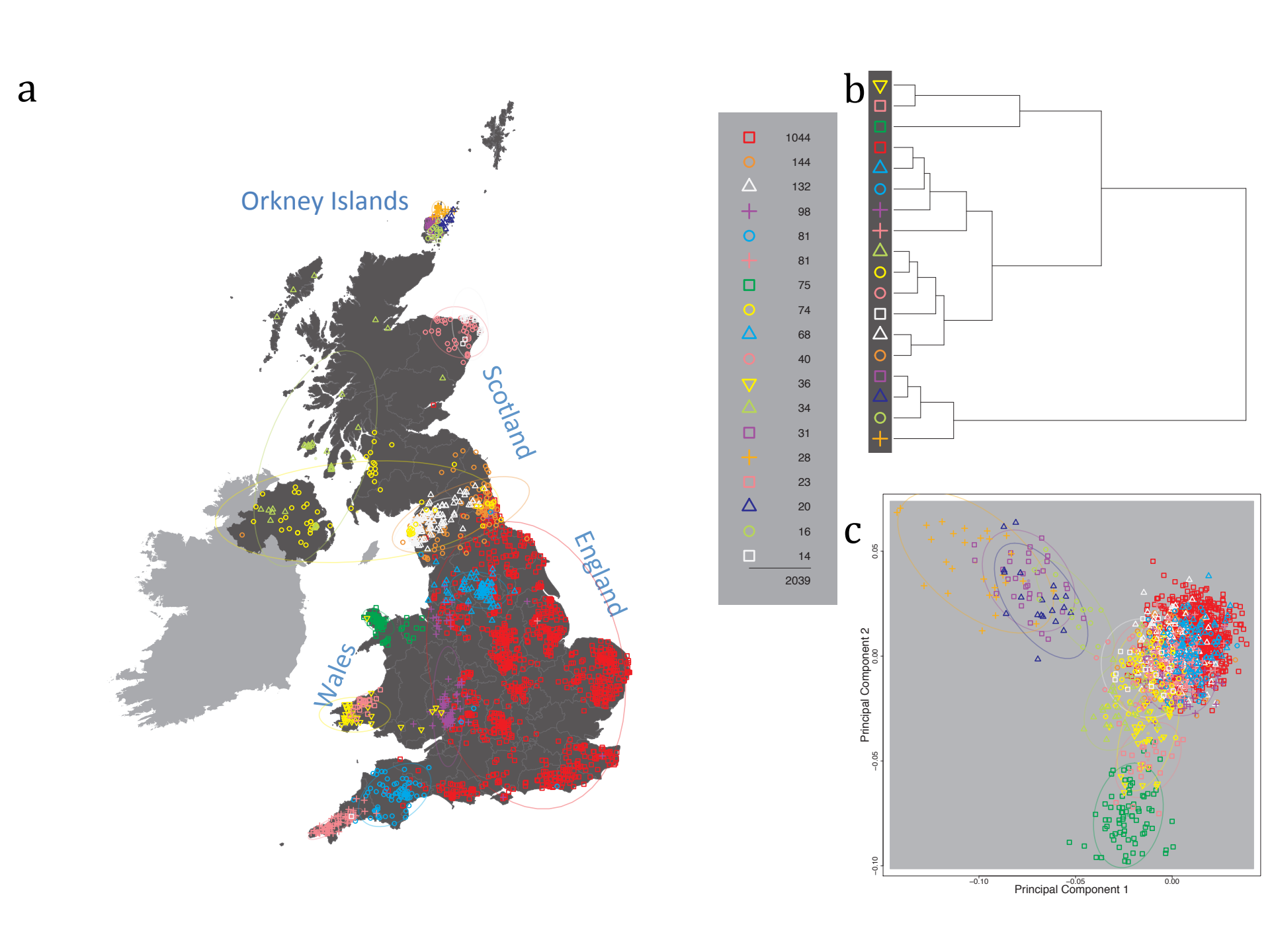

a
Orkney Islands
Scotland
England
Wales
1112
276
108
98
81
81
75
59
51
40
28
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

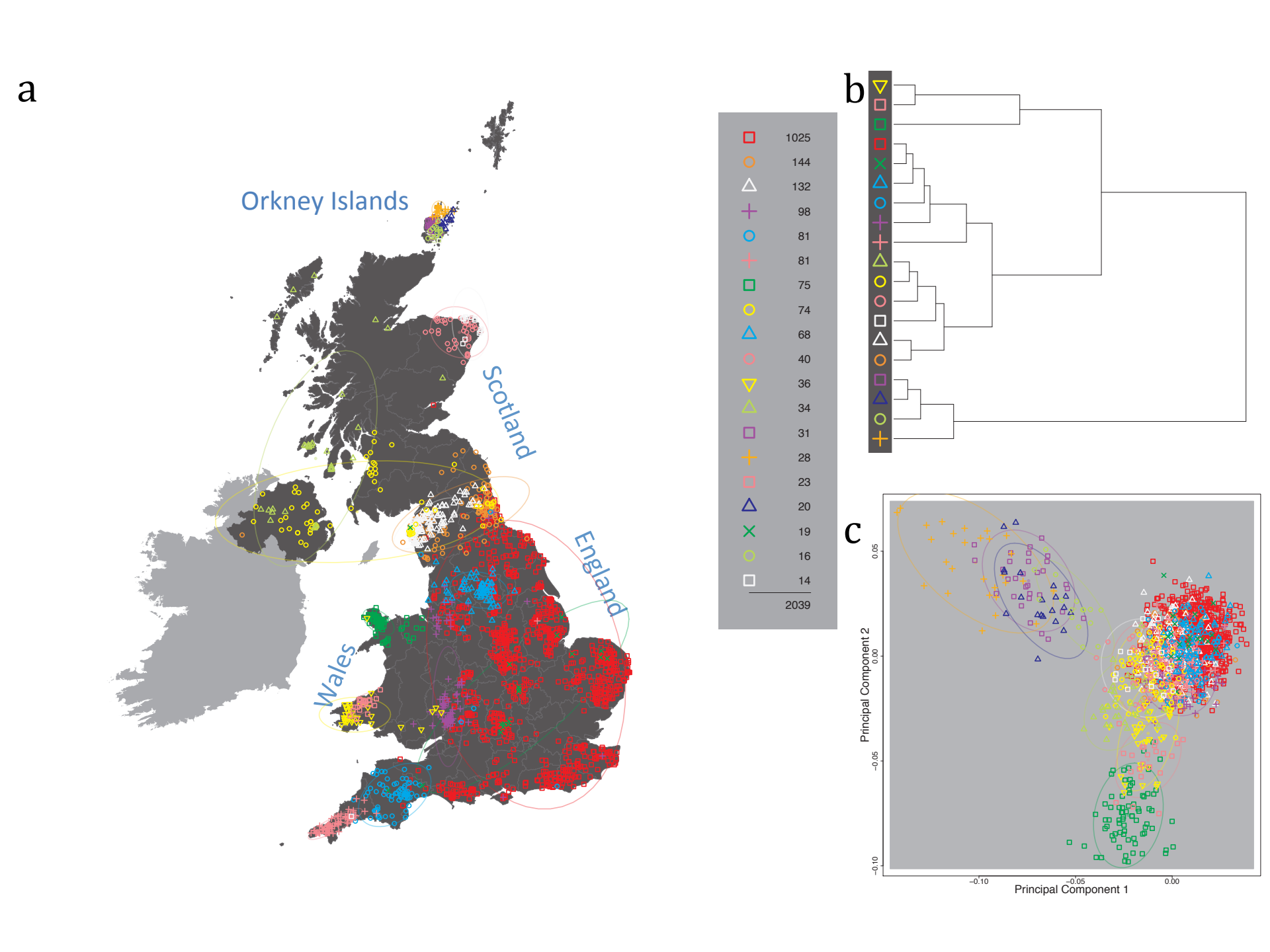

a
Orkney Islands
Scotland
England
Wales
1112
276
108
98
81
81
75
51
40
36
28
23
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

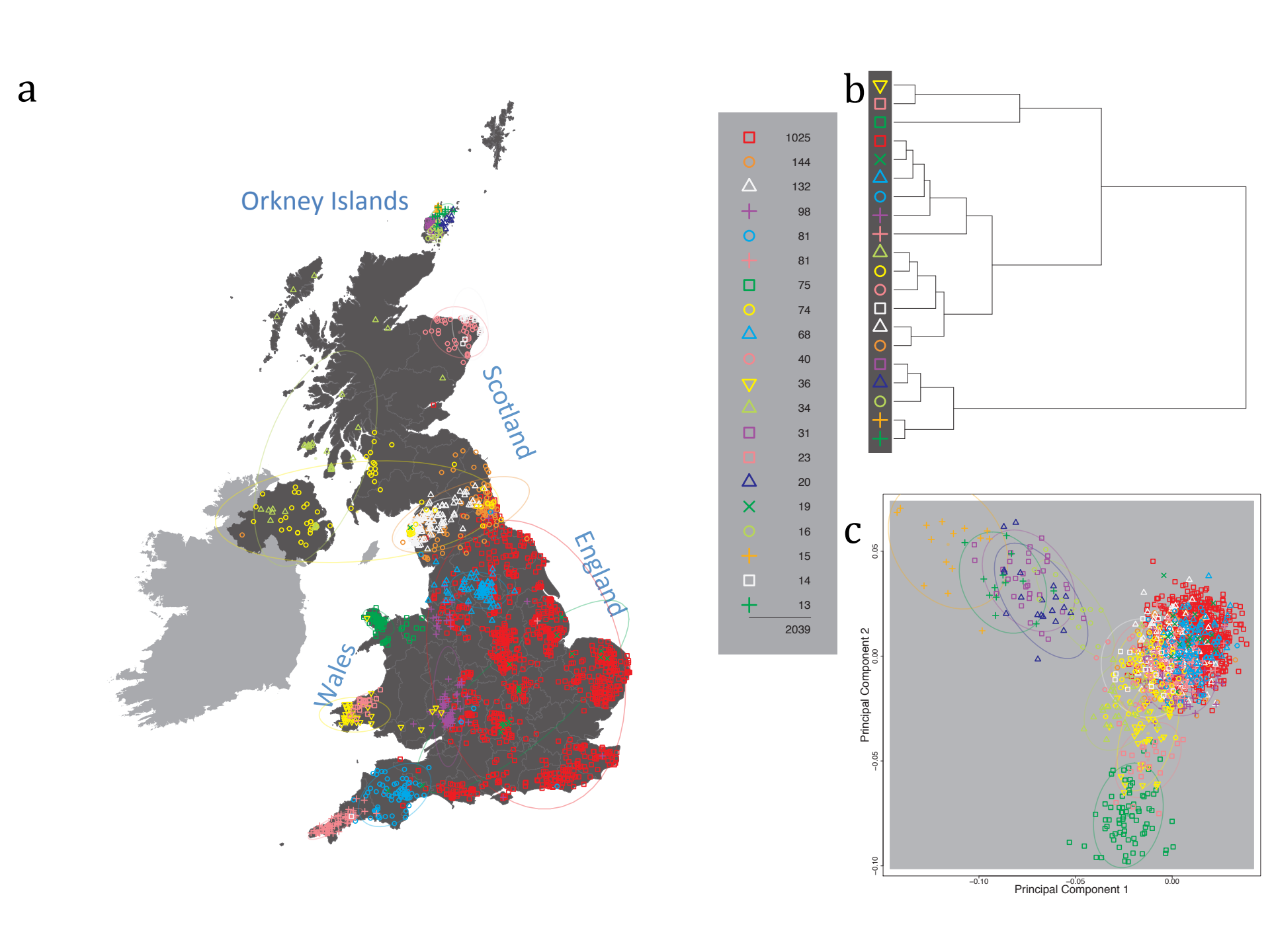

a
Orkney Islands
Scotland
England
Wales
1044
276
108
98
81
81
75
68
51
40
36
28
23
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

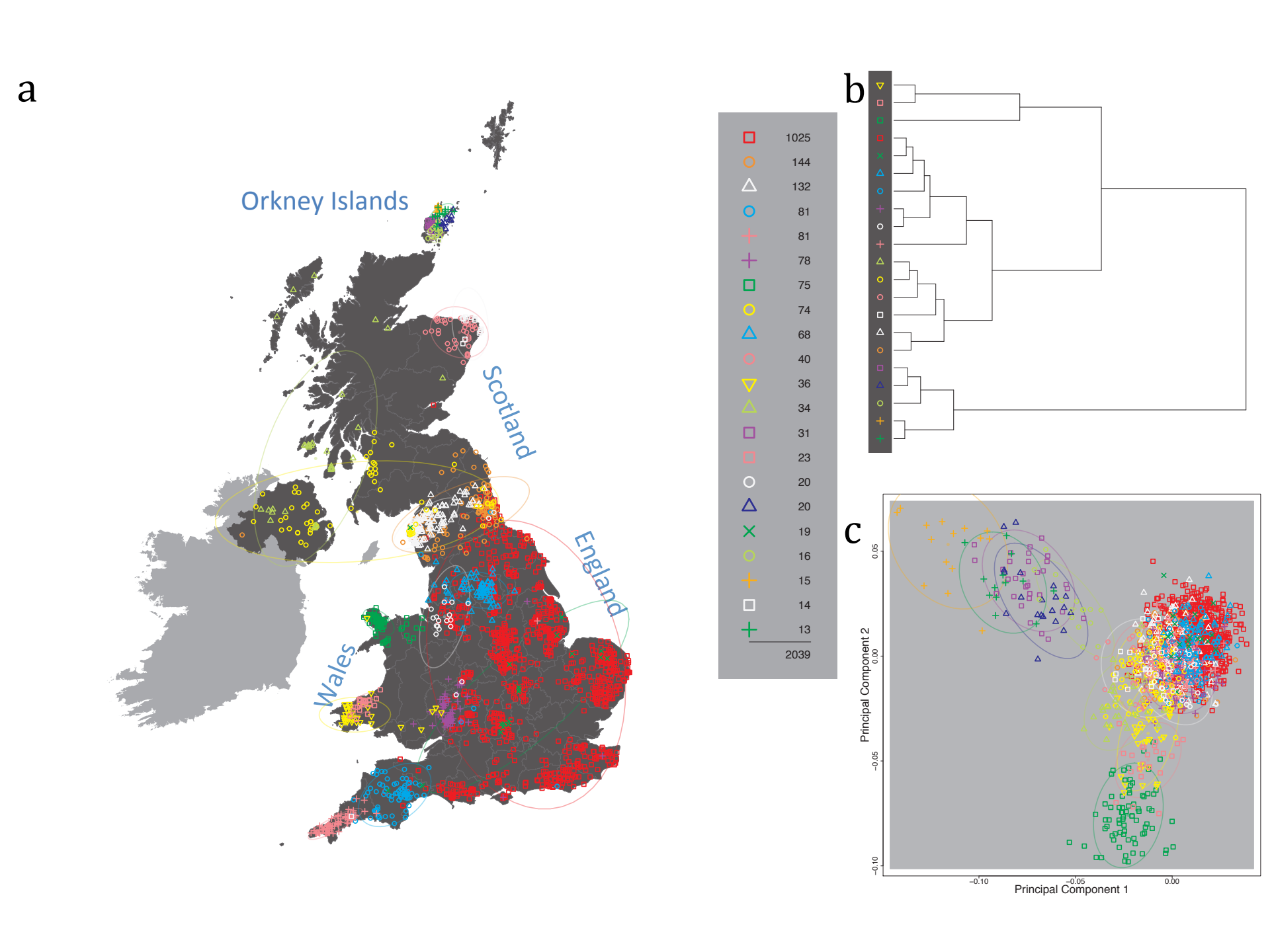

a
Orkney Islands
Scotland
England
Wales
1044
144
132
108
98
81
81
75
68
51
40
36
28
23
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

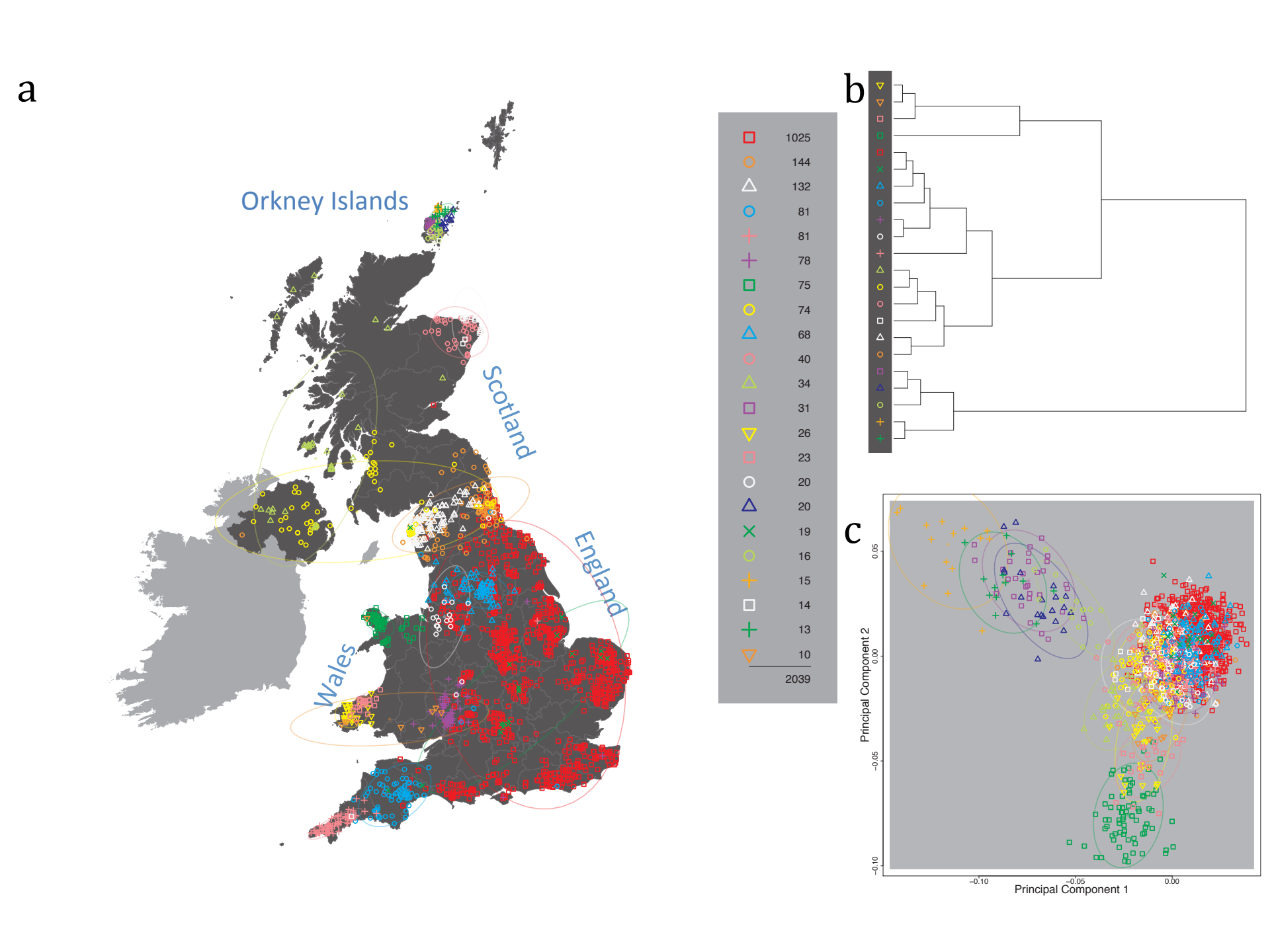

a
Orkney Islands
Scotland
England
Wales
1044
144
132
98
81
81
75
74
68
51
40
36
34
28
23
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

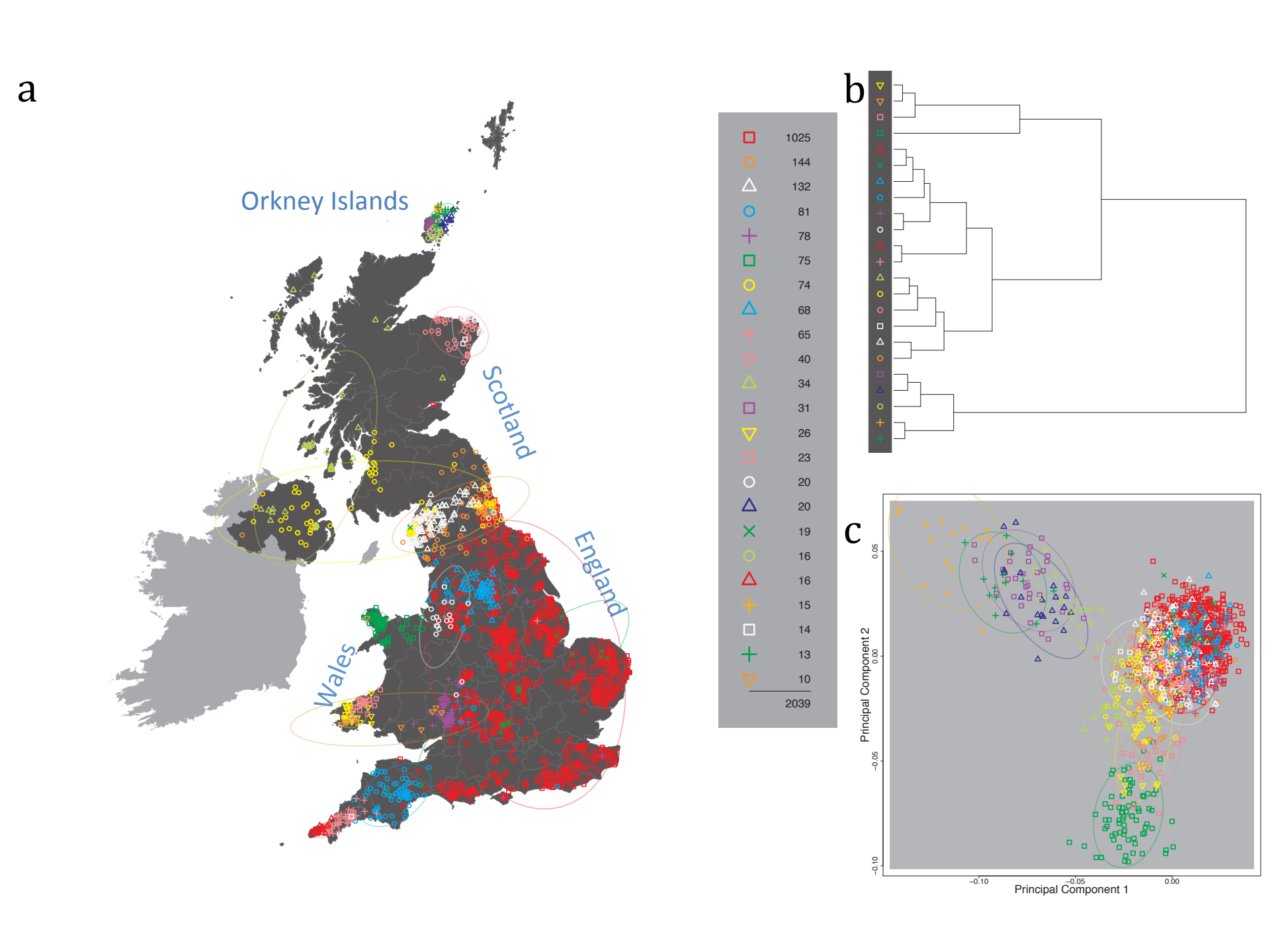

a
Orkney Islands
Scotland
England
Wales
1044
144
132
98
81
81
75
74
68
40
36
34
31
28
23
20
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

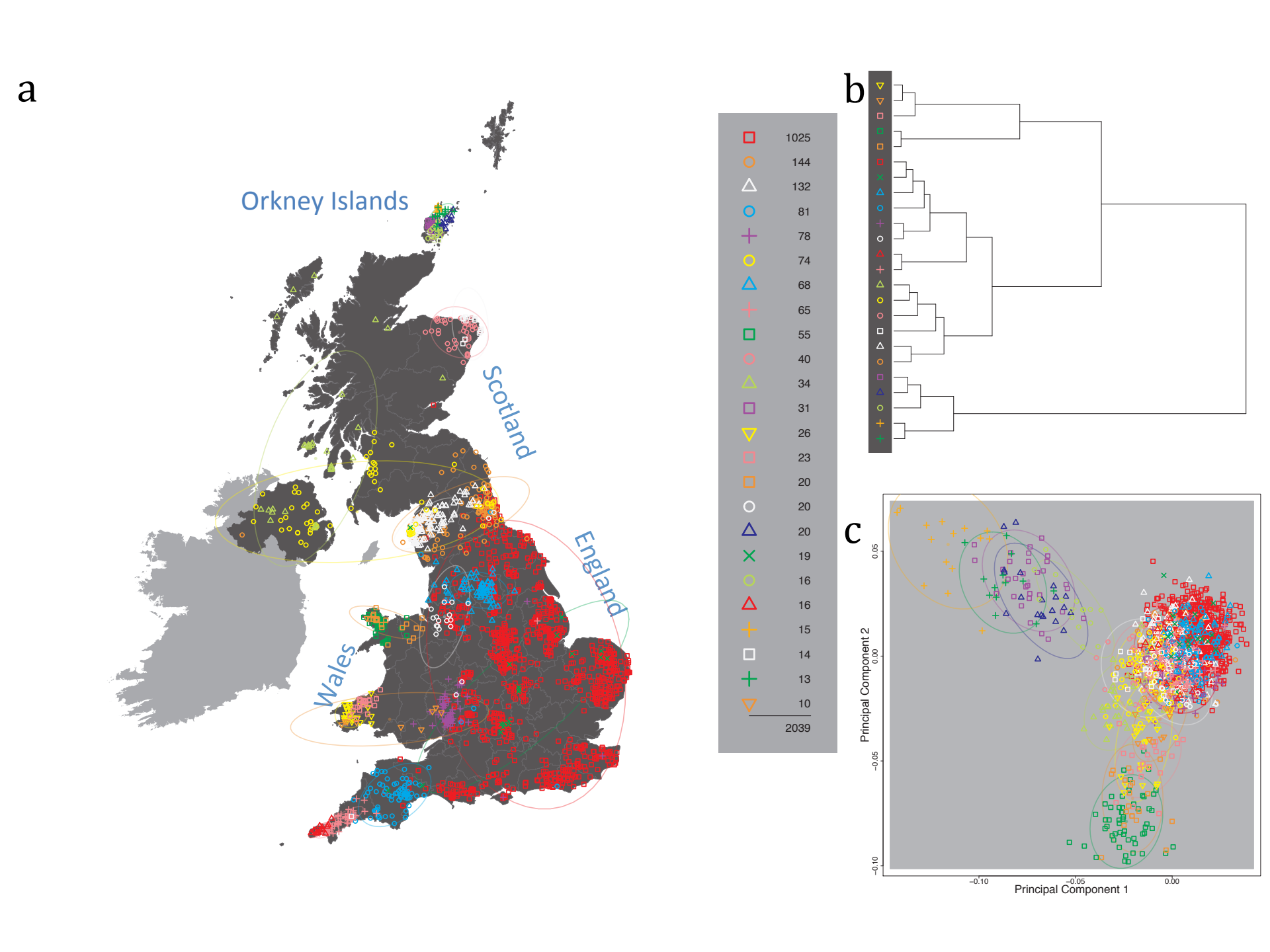

a
Orkney Islands
Scotland
England
Wales
1025
144
132
98
81
81
75
74
68
40
36
34
31
28
23
20
19
16
14
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00

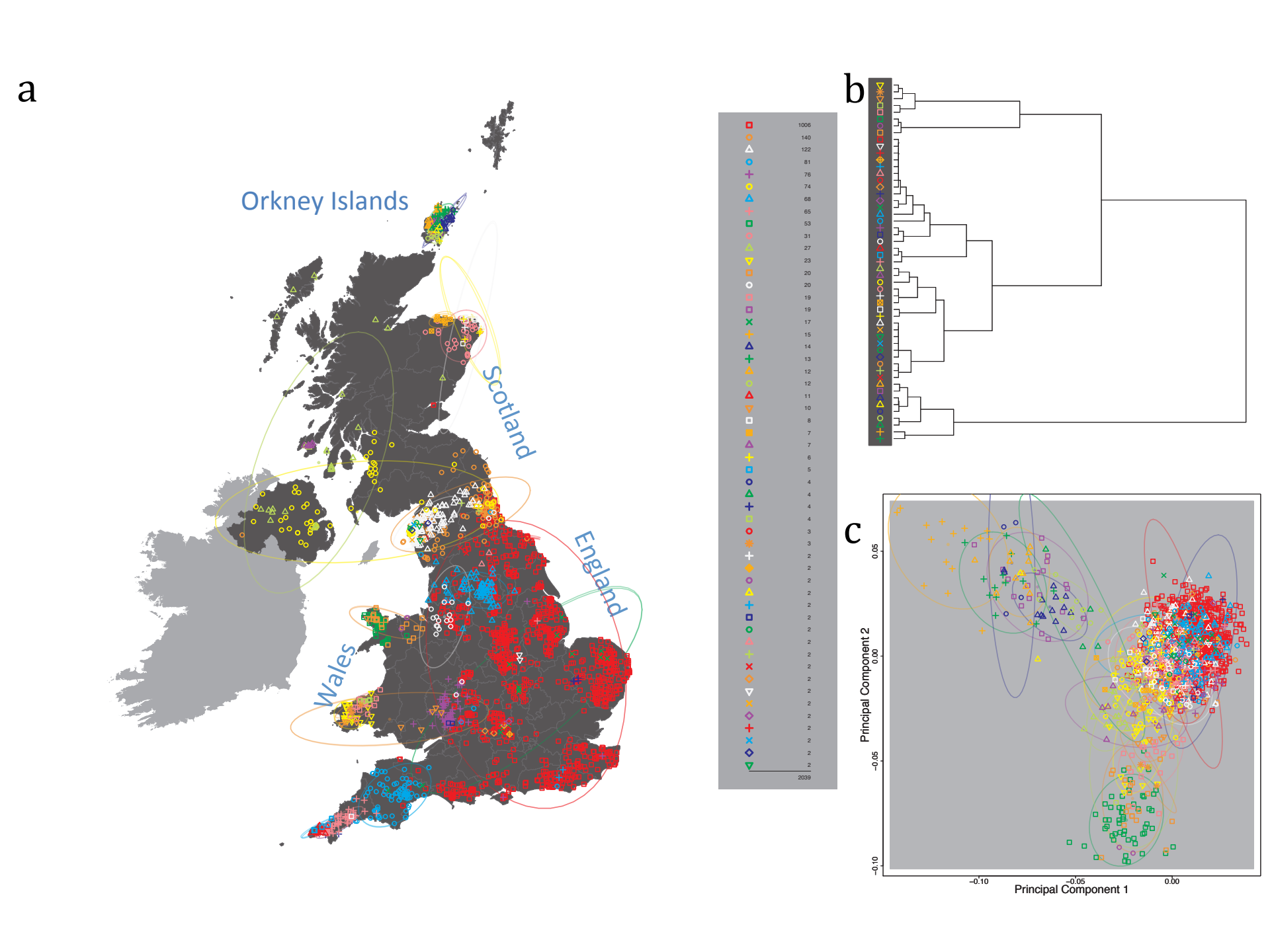

a
Orkney Islands
Scotland
England
Wales
1025
144
132
98
81
81
75
74
68
40
36
34
31
23
20
19
16
15
14
13
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00


a
Orkney Islands
Scotland
England
Wales
1025
144
132
81
81
78
75
74
68
40
36
34
31
23
20
20
19
16
15
14
13
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00


a
Orkney Islands
Scotland
England
Wales
1025
144
132
81
81
78
75
74
68
40
34
31
26
23
20
20
19
16
15
14
13
10
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00


a
Orkney Islands
Scotland
England
Wales
1025
144
132
81
78
75
74
68
65
40
34
31
26
23
20
20
19
16
16
15
14
13
10
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00


a
Orkney Islands
Scotland
England
Wales
1025
144
132
81
78
74
68
65
55
40
34
31
26
23
20
20
20
19
16
16
15
14
13
10
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00


a
Orkney Islands
Scotland
England
Wales
1006
140
122
81
76
74
68
65
53
31
27
23
20
20
19
19
17
15
14
13
12
12
11
10
8
7
7
6
5
4
4
4
4
3
3
2
2
2
2
2
2
2
2
2
2
2
2
2
2
2
2
2039
b
c
Principal Component 2
Principal Component 1
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00