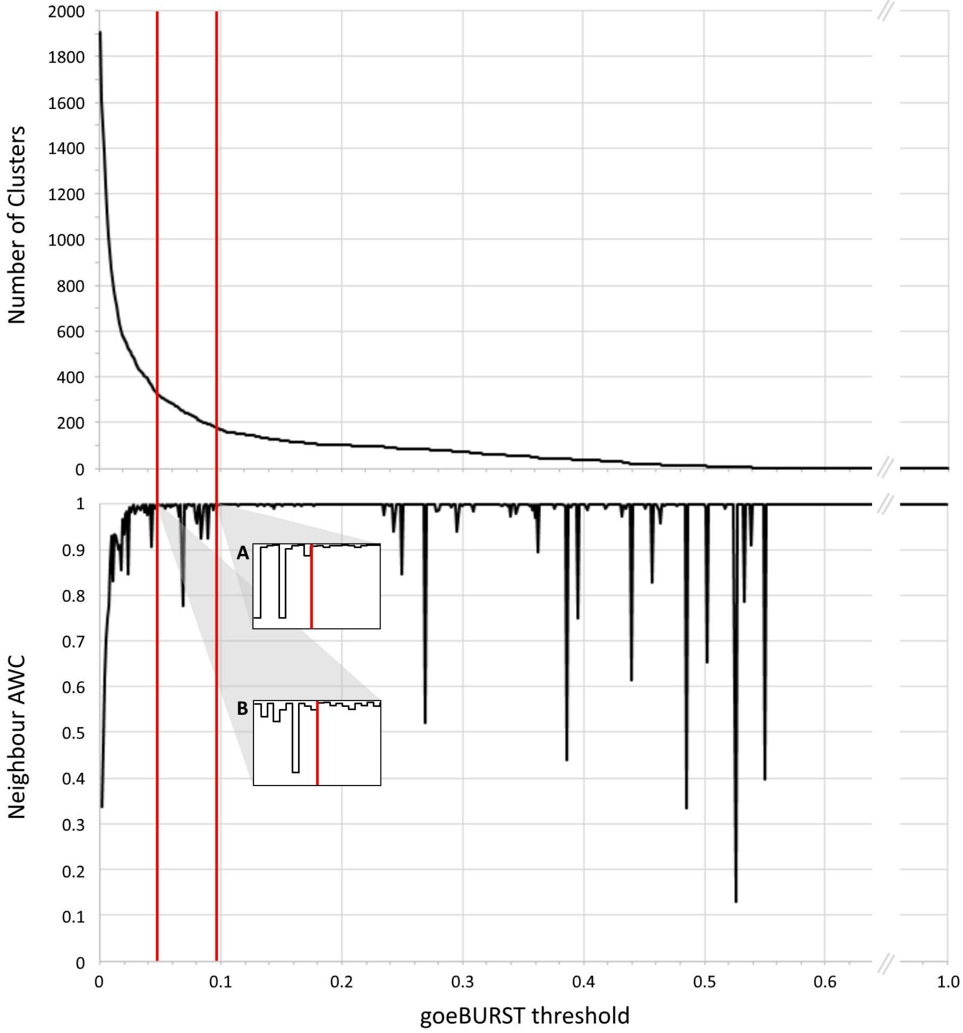

**Fig. S1** – Relationship between cluster stability (assessed by Neighbourhood Adjusted Wallace Coefficient) and goeBURST threshold based on the MScgMLST schema for all 3791 *N. gonorrhoeae* isolates analysed. **A** – Second point of cluster stability, corresponding to a goeBURST threshold of 0.096 (i.e., 79 allelic differences). **B** – First point of cluster stability, corresponding to a goeBURST threshold of 0.048 (i.e., 40 allelic differences).

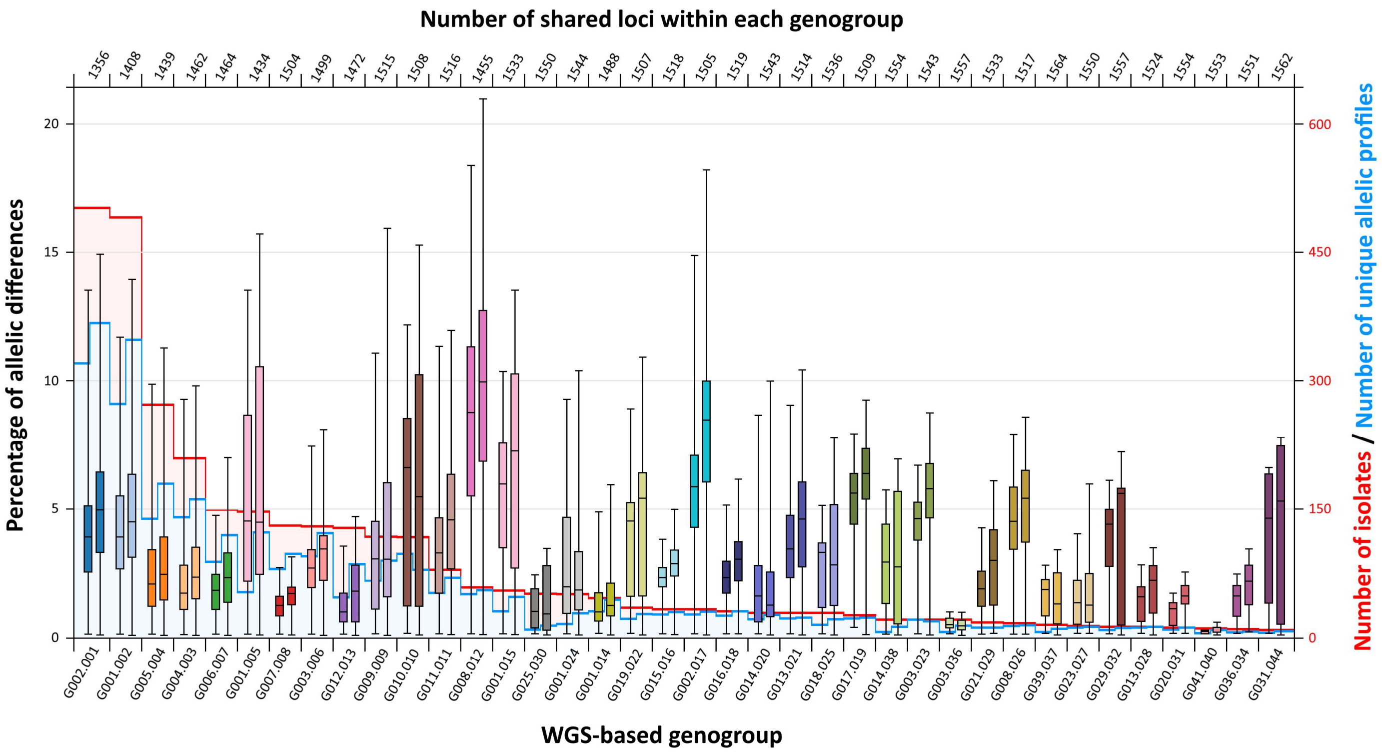

**Fig. S2** – Distribution of the percentage of pairwise allelic differences (AD) between isolates of all low-level WGS-based genogroups represented by at least 10 isolates. Data for each WGS-based genogroup is presented by pair of box plots with the same colour. Left and right box plots refers to AD distribution using the MScgMLST loci schema or the maximum shared loci of the PubMLST cgMLST schema within the genogroup, respectively. The respective total number of loci shared is presented on top of the graph. Red line refers to the number of isolates that constitute each low-level WGS-based genogroup, while the blue line refers to the number of unique allelic profiles detected within each genogroup using the two approaches.

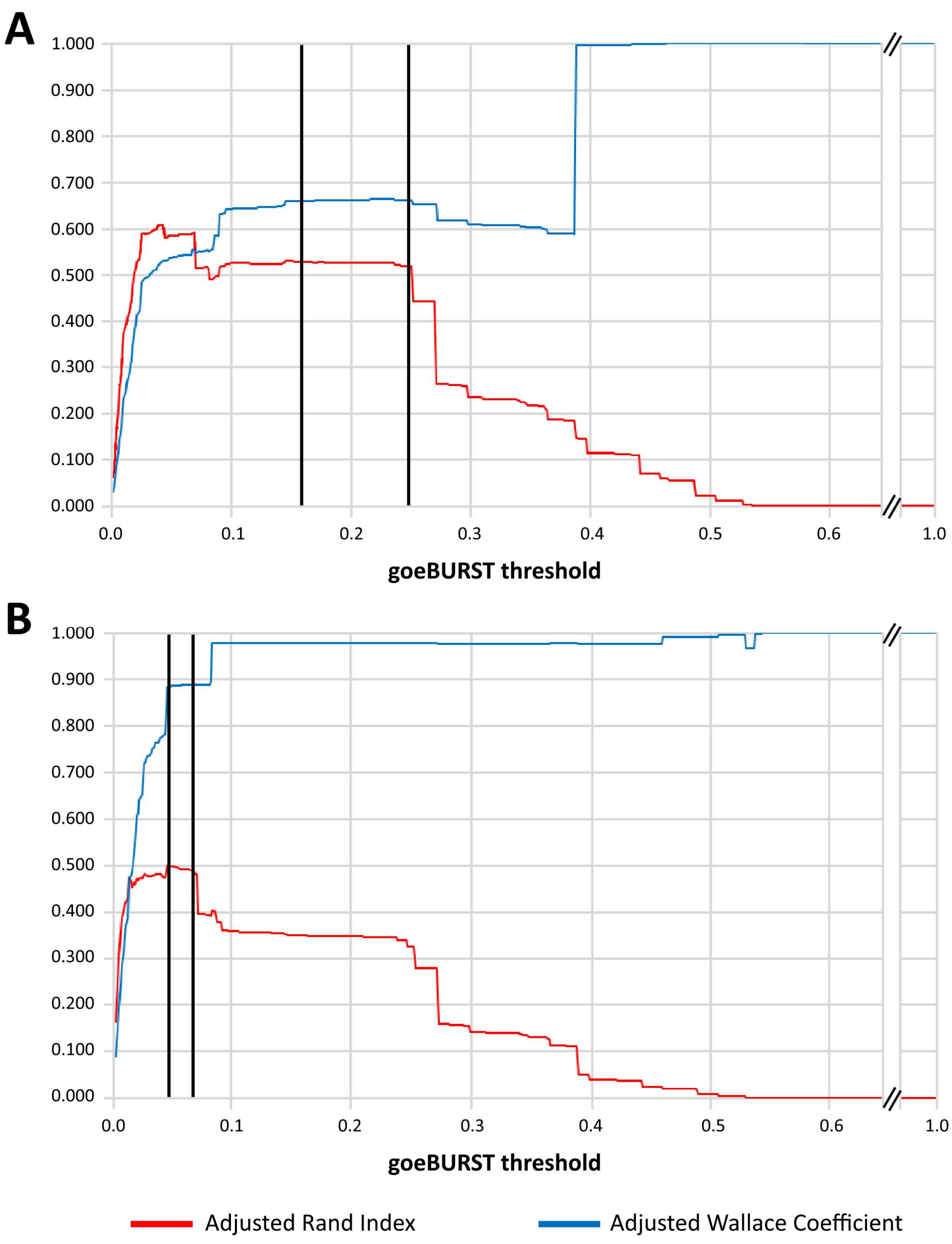

**Fig. S3** – Cluster congruence (Adjusted Rand Index) and agreement (Adjusted Wallace Coefficient) between traditional (**A**) MLST and (**B**) NG-MAST typing and goeBURST clustering at all thresholds (based on the MScgMLST schema). Adjusted Wallace Coefficient values are directed from traditional typing to goeBURST clustering. Black lines outline the threshold ranges where higher congruence and agreement was simultaneously observed.

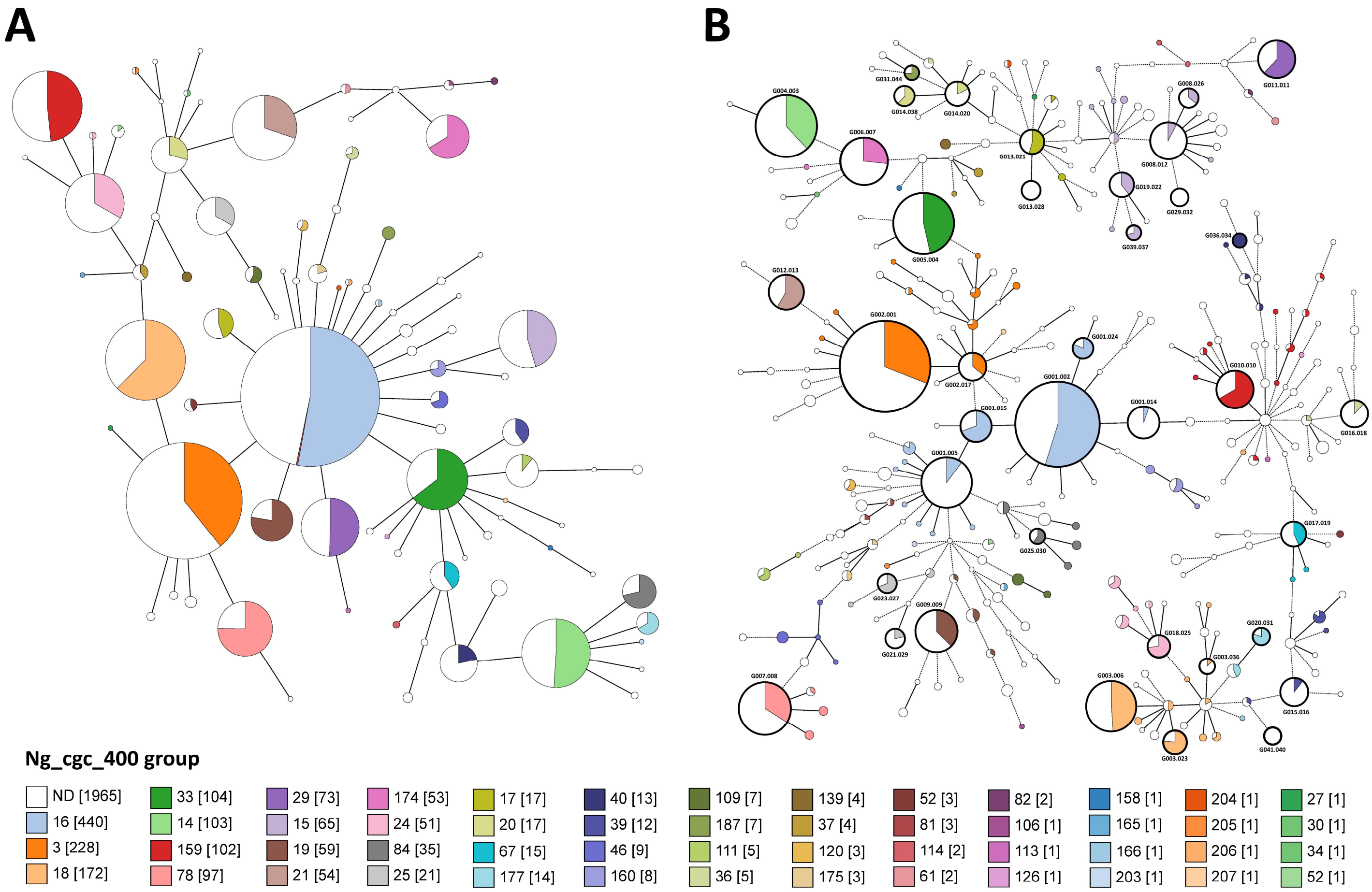

**Fig. S4 –** Overlap between *N. gonorrhoeae* WGS-based genogroups defined in this study and pubMLST core-genome groups. Minimum spanning trees (MST) were constructed based on allelic diversity found among the 822 genes shared by 100% of the isolates and nodes are coloured according to Ng_cgc_400 groups [28], for the sub-set of isolates used in both studies (N = 1830). **A** – MST nodes were collapsed at the pubMLST threshold (~24% AD), i.e., 199 AD in 822 loci MScgMLST scheme which correspond to 400 AD in 1649 loci pubMLST scheme, illustrating the huge clustering overlap between both strategies. **B** – MST nodes were collapsed at the low-level WGS-genogroup threshold (4.87%; 40 AD), with straight and dotted lines reflecting nodes linked with the AD below and above the threshold applied for the high-level (9.61%; 79 AD), illustrating that the stable threshold defined in the present study provide a higher discriminatory resolution than Ng_cgc_400. Numbers in parenthesis refer to the number of isolates comprising each Ng_cgc_400. MSTs were generated using the GrapeTree v1.5.0 software [56].

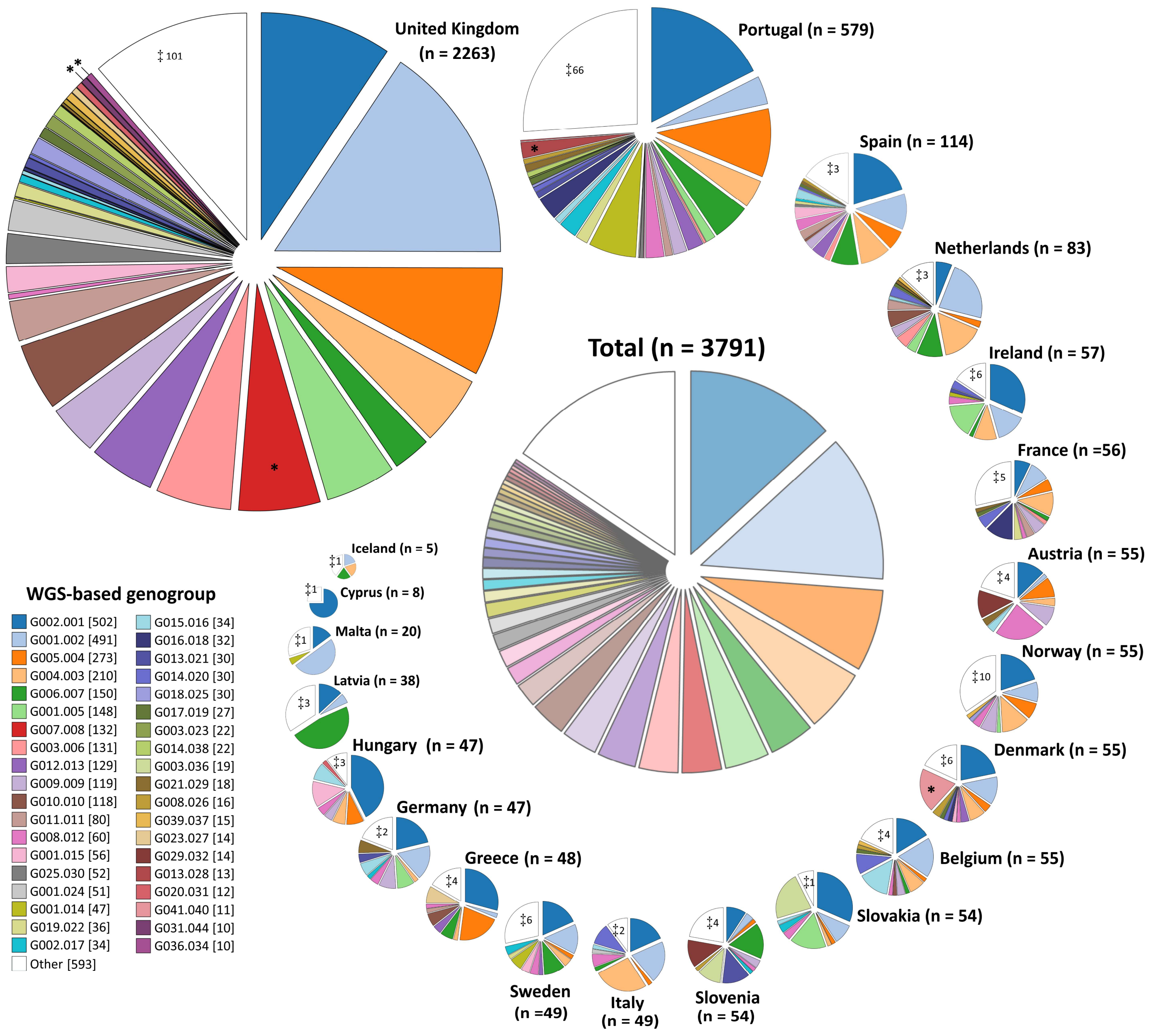

**Fig. S5** – Distribution of low-level WGS-based genogroups by country. Only genogroups with at least 10 isolates are coloured. Central pie chart refers to the total distribution of genogroups within the analysed dataset. *WGS-based genogroups only found in one country, represented by at least 10 isolates. ‡WGS-based genogroups only found in one country, represented by less than 10 isolates (adjacent number refers to the number of distinct genogroups).