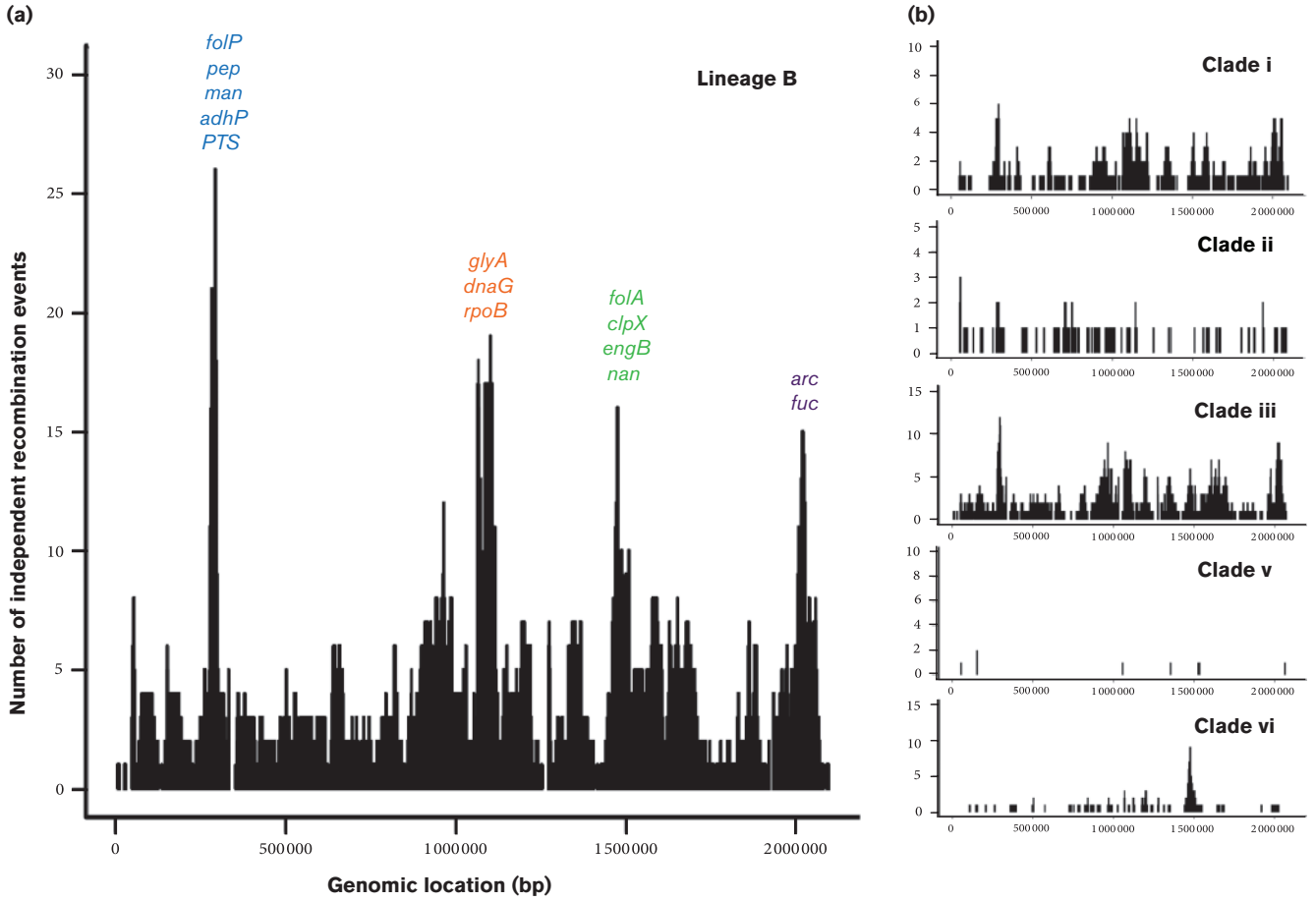

Fig. 4. (a). Genomic regions recombination in the 376 lineage B isolates. The genes subject to the highest number of independent recombination events are named. (b). Genomic regions under recombination within the five lineage B Africa clades. *folP*, dihydropteroate synthase; *pep*, aminopeptidase; *man*, mannose PTS system component; *adhP*, alcholol dehydrogenase; *PTS*; phosphotransferase system protein; *glyA*, serine hydroxymethyltransferase; *dnaG*, DNA primase; *rpoB*, RNA polymerase beta subunit; *folA*, dihyrdofolate reductase; *clpX*, CLP protease; *engB*, STP binding protein; *nan*, neuraminidase; *arc*, arginine deiminase; *fuc*, fuculose kinase.

the Malawi and Mozambique isolates in clade vi and a subset of North African isolates (clade iii). A homologue of PpiA reduces phagocytosis in *Streptococcus mutans* (Iyer *et al.*, 2001). Three genes from a phosphotransferase system (PTS) were present in all of the isolates in the North African clades (i and ii). PTS plays a key role in pneumococcal colonization (Mukouhara *et al.*, 2011). The distribution of these accessory genes may reflect host-specific selective pressures between regions and explain differences in disease severity between these regions.

As discussed earlier, the lineage B phylogeny (Fig. 2) indicates that there were two distinct, genetically unrelated groups of isolates from the Gambia, ST618 and ST3081. In 2006, ST3081 replaced ST618 as the dominant cause of IPD in the Gambia. MLST does not have the resolution to distinguish the genes exclusive to ST3081, which have driven sequence type (ST) replacement. To look at large accessory elements, rather than absence/presence of individual genes, we took a semi-manual approach to the creation of the Gambia accessory genome (see Methods) (Fig. 8). Accessory regions present in all ST3081 isolates but absent in ST618 isolates included a 6.2 kb phage element, encoding an integrase site-specific recombinase (XerD), a cell division protein, FtsK/SpoIIIE, and a fucose utilization operon first characterized in serotype 3 (SP3-BS71) (Higgins *et al.*, 2009). Unique to all of the ST618 isolates was a type 2 fucose utilization operon, first characterized in *S. pneumoniae* TIGR4 (Chan *et al.*, 2003) (Fig. S8). Fucose utilization operons have been shown to be essential to pneumococcal virulence in models of pneumonia and otitis media (Embry *et al.*, 2007). It is possible that the SP3-BS71-type operon found in ST3081 allows ST3081 to outcompete ST618 isolates, which possess the type 2 operon. Of the remaining proteins unique to ST3081, XerD has previously been implicated in pneumococcal virulence (Chalker *et al.*, 2000). FtsK has not previously been implicated in pneumococcal virulence; however, it interacts with Xer proteins (Le