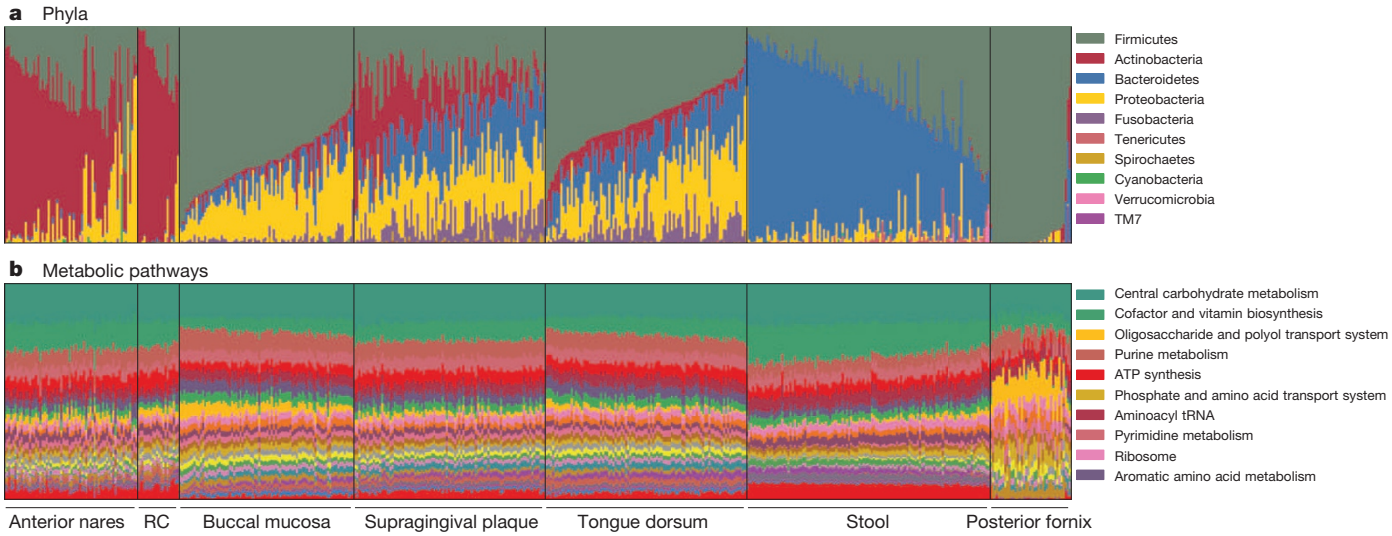

Figure 2 | Carriage of microbial taxa varies while metabolic pathways remain stable within a healthy population. **a**, **b**, Vertical bars represent microbiome samples by body habitat in the seven locations with both shotgun and 16S data; bars indicate relative abundances colored by microbial phyla from binned OTUs (**a**) and metabolic modules (**b**). Legend indicates most abundant phyla/pathways by average within one or more body habitats; RC, retroauricular crease. A plurality of most communities' memberships consists of a single dominant phylum (and often genus; see Supplementary Fig. 2), but this is universal neither to all body habitats nor to all individuals. Conversely, most metabolic pathways are evenly distributed and prevalent across both individuals and body habitats.

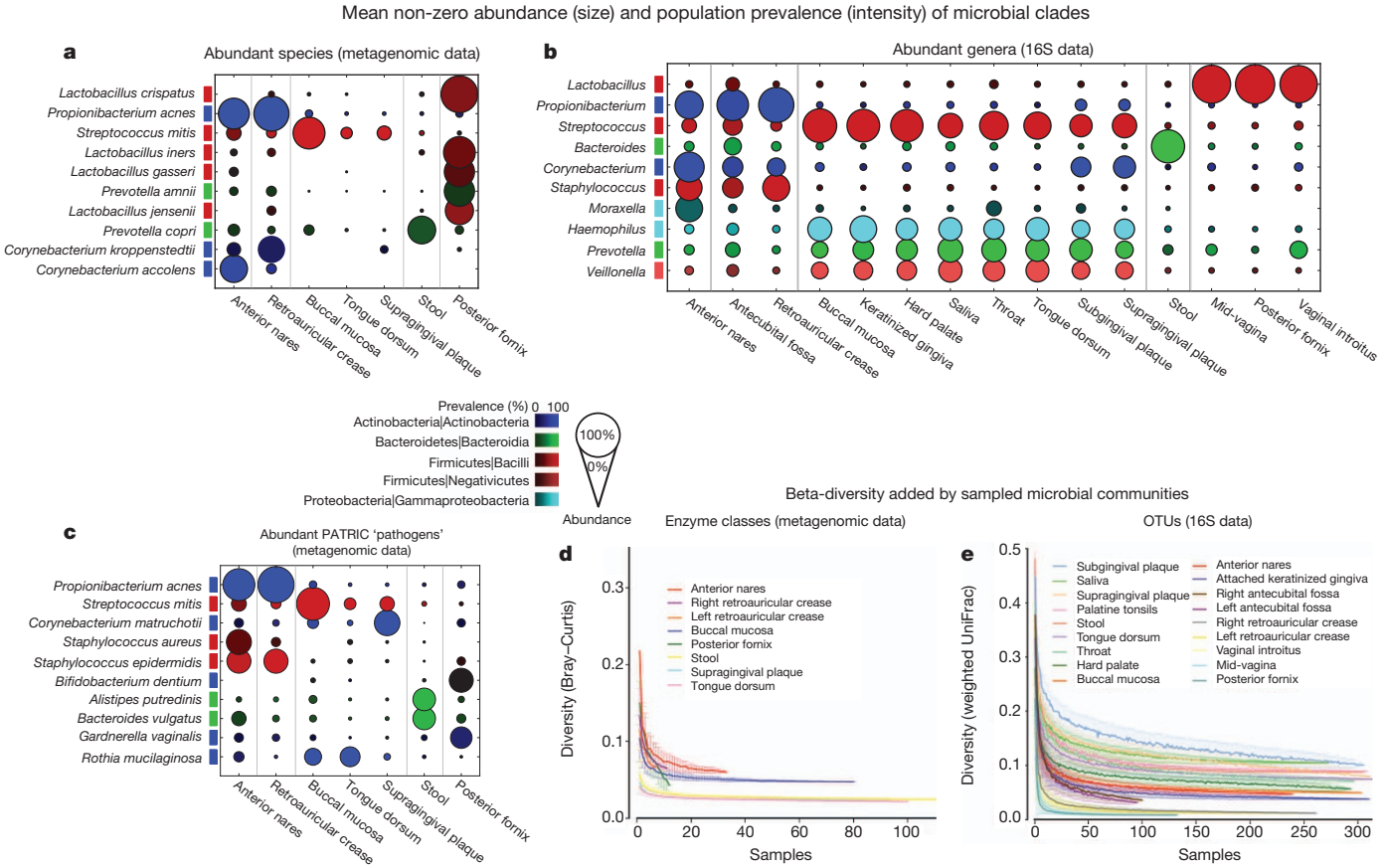

Figure 3 | Abundant taxa in the human microbiome that have been metagenomically and taxonomically well defined in the HMP population. **a**–**c**, Prevalence (intensity, colour denoting phylum/class) and abundance when present (size) of clades in the healthy microbiome. The most abundant metagenomically-identified species (**a**), 16S-identified genera (**b**) and PATRIC[12] pathogens (metagenomic) (**c**) are shown. **d**, **e**, The population size and sequencing depths of the HMP have well defined the microbiome at all assayed body sites, as assessed by saturation of added community metabolic configurations (rarefaction of minimum Bray–Curtis beta-diversity of metagenomic enzyme class abundances to nearest neighbour, inter-quartile range over 100 samples) (**d**) and phylogenetic configurations (minimum 16S OTU weighted UniFrac distance to nearest neighbour) (**e**).

Carriage of specific microbes

Inter-individual variation in the microbiome proved to be specific, functionally relevant and personalized. One example of this is illustrated by the *Streptococcus* spp. of the oral cavity. The genus dominates the oropharynx[16], with different species abundant within each sampled body habitat (see http://hmpdacc.org/HMSMCP) and, even at the species level, marked differences in carriage within each habitat among individuals (Fig. 4a). As the ratio of pan- to core-genomes is high in many human-associated microbes[17], this variation in abundance could be due to selective pressures acting on pathways differentially present among *Streptococcus* species or strains (Fig. 4b). Indeed, we observed extensive strain-level genomic variation within microbial species in this population, enriched for host-specific structural variants around genomic islands (Fig. 4c). Even with respect to the single *Streptococcus mitis* strain B6, gene losses associated with these events were common, for example differentially eliminating *S. mitis* carriage of the V-type ATPase or choline binding proteins cbp6 and cbp12 among subsets of the host population (Fig. 4d). These losses were easily observable by comparison to reference isolate genomes, and these initial findings indicate that microbial strain- and host-specific gene gains and polymorphisms may be similarly ubiquitous.

Other examples of functionally relevant inter-individual variation at the species and strain levels occurred throughout the microbiome. In the gut, *Bacteroides fragilis* has been shown to prime T-cell responses in animal models via the capsular polysaccharide A[18], and in the HMP stool samples this taxon was carried at a level of at least 0.1% in 16% of samples (over 1% abundance in 3%). *Bacteroides thetaiotaomicron* has been studied for its effect on host gastrointestinal metabolism[19] and was likewise common at 46% prevalence. On the skin, *S. aureus*, of particular interest as the cause of methicillin-resistant

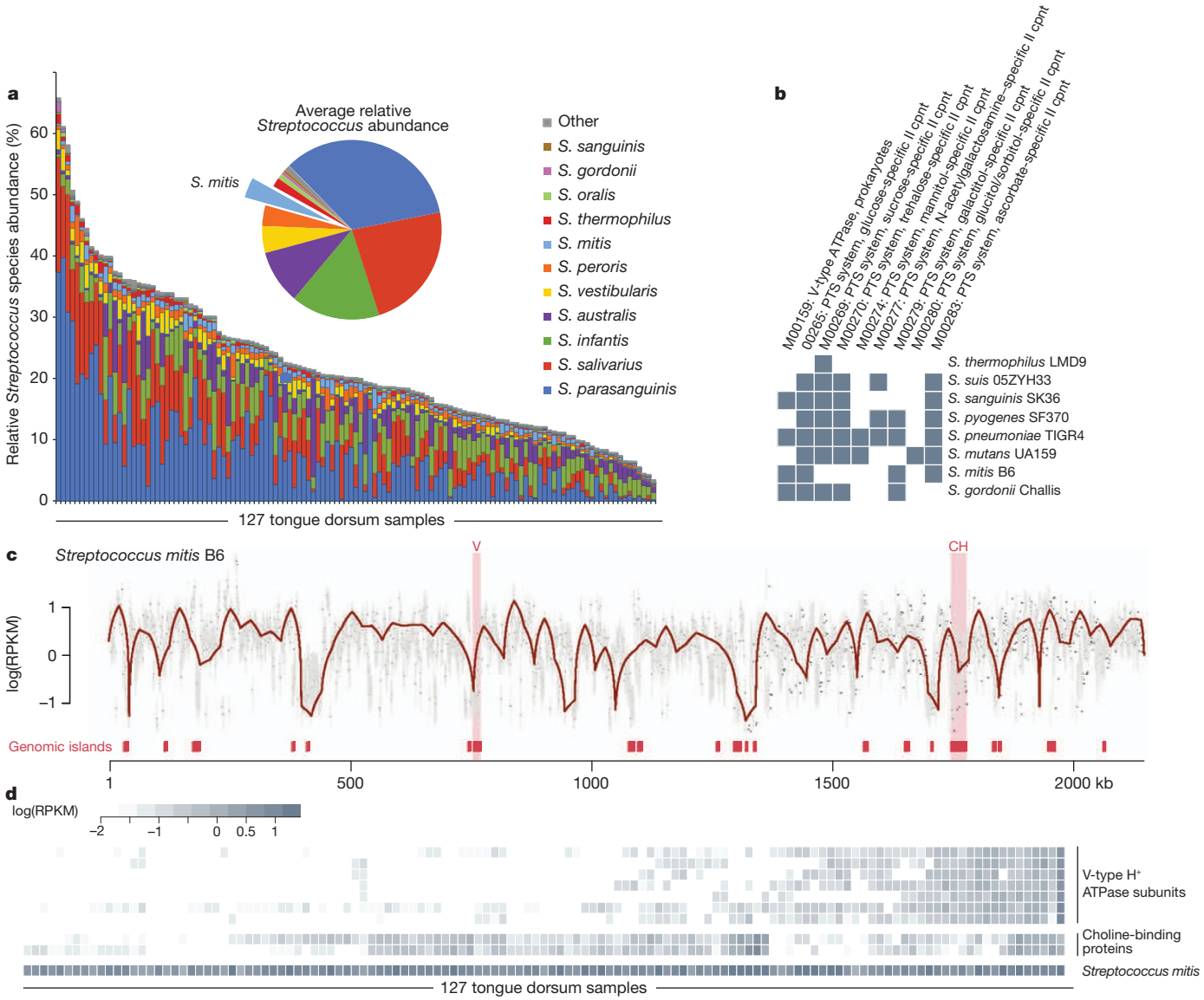

Figure 4 | Microbial carriage varies between subjects down to the species and strain level. Metagenomic reads from 127 tongue samples spanning 90 subjects were processed with MetaPhlAn to determine relative abundances for each species. **a**, Relative abundances of 11 distinct *Streptococcus* spp. In addition to variation in broader clades (see Fig. 2), individual species within a single habitat demonstrate a wide range of compositional variation. Inset illustrates average tongue sample composition. **b**, Metabolic modules present/absent (grey/white) in KEGG[24] reference genomes of tongue streptococci denote selected areas of strain-specific functional differentiation. cpnt, component. **c**, Comparative genomic coverage for the single *Streptococcus mitis* B6 strain. Grey dots are median reads per kilobase per million reads (RPKM) for 1-kb windows, grey bars are the 25th to 75th percentiles across all samples, red line the LOWESS-smoothed average. Red bars at the bottom highlight predicted genomic islands[27]. Large, discrete, and highly variable islands are commonly under-represented. **d**, Two islands are highlighted, V (V-type H^+ ATPase subunits I, K, E, C, F, A and B) and CH (choline-binding proteins cbp6 and cbp12), indicating functional cohesion of strain-specific gene loss within individual human hosts.