ACKNOWLEDGMENTS
We thank our laboratory members for their insights and helpful discussions. This work was supported by the National Research Foundation of Korea (grant 2011-0013901 to J.-B.L.) and the US National Institutes of Health (grant AI099854 to K.E.Y.; grants GM080176 and CA67007 to R.F.).

AUTHOR CONTRIBUTIONS
G.S., J.L., J.-B.L., K.E.Y. and R.F. designed the experiments; G.S., J.L. and R.F. analyzed the data; M.A.L. helped with the PCD purification and performed PFV integration analysis; J.M.-L. prepared the D-loop DNA substrate; J.H. helped with the SM analysis; all authors participated in critical discussions and writing of the paper.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Gayan Senavirathne[1,5], Jiaquan Liu[1,5], Miguel A Lopez Jr[1], Jeungphill Hanne[1], Juana Martin-Lopez[1], Jong-Bong Lee[2,3], Kristine E Yoder[1] & Richard Fishel[1,4]

[1]Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University Medical Center, Columbus, Ohio, USA. [2]Department of Physics, Pohang University of Science and Technology, Pohang, Korea. [3]School of Interdisciplinary Bioscience and Bioengineering, Pohang, Korea. [4]Physics Department, The Ohio State University, Columbus, Ohio, USA. [5]These authors contributed equally to this work.
e-mail: fishel.7@osu.edu

1. Rasnik, I., McKinney, S.A. & Ha, T. *Nat. Methods* **3**, 891–893 (2006).
2. Shi, X., Lim, J. & Ha, T. *Anal. Chem.* **82**, 6132–6138 (2010).
3. Aitken, C.E., Marshall, R.A. & Puglisi, J.D. *Biophys. J.* **94**, 1826–1835 (2008).
4. Ohlendorf, D.H., Lipscomb, J.D. & Weber, P.C. *Nature* **336**, 403–405 (1988).
5. Juette, M.F. *et al. Curr. Opin. Chem. Biol.* **20**, 103–111 (2014).

MetaPhlAn2 for enhanced metagenomic taxonomic profiling

To the Editor: MetaPhlAn (metagenomic phylogenetic analysis)[1] is a method for characterizing the taxonomic profiles of whole-metagenome shotgun (WMS) samples that has been used successfully in large-scale microbial community studies[2,3]. This work complements the original species-level profiling method with a system for eukaryotic and viral quantitation, strain-level identification and strain tracking. These and other extensions make the MetaPhlAn2 computational package (http://segatalab.cibio.unitn.it/tools/metaphlan2/ and **Supplementary Software**) an efficient tool for mining WMS samples.

Our method infers the presence and read coverage of clade-specific markers to unequivocally detect the taxonomic clades present in a microbiome sample and estimate their relative abundance[1]. MetaPhlAn2 includes an expanded set of ~1 million markers (184 ± 45 for each bacterial species) from >7,500 species (**Supplementary Tables 1–3**), based on the approximately tenfold increase in the number of sequenced genomes in the past 2 years. Subspecies markers enable strain-level analyses, and quasi-markers improve accuracy and allow the detection of viruses and eukaryotic microbes (a full list of additions is provided in **Supplementary Notes 1–3** and **Supplementary Fig. 1**).

We validated MetaPhlAn2 using 24 synthetic metagenomes comprising 656 million reads and 1,295 species (**Supplementary Note 4** and **Supplementary Table 4**). MetaPhlAn2 proved more accurate (average correlation: 0.95 ± 0.05) than mOTU[4] and Kraken[5] (0.80 ± 0.21 and 0.75 ± 0.22, respectively) (**Fig. 1a**, **Supplementary Figs. 2–9** and **Supplementary Tables 5–11**), with fewer false positives (an average of 10, compared with 22 and 23 for mOTU and Kraken, respectively) and false negatives (an average of 12, compared with 27 for the other two methods), even when including genomes that were absent from the reference database (**Supplementary Note 4**). With the adoption of the BowTie2 fast mapper and support for parallelism, MetaPhlAn2 is more than ten times faster than MetaPhlAn, and its speed is comparable to that of other tested approaches (**Supplementary Fig. 10**).

We applied MetaPhlAn2 to four elbow-skin samples that we sequenced from three subjects (**Fig. 1b**, **Supplementary Note 5** and **Supplementary Table 12**). Our data showed that *Propionibacterium acnes* and *Staphylococcus epidermidis* dominated these sites, in agreement with expected genus-level results[6], while providing species-level resolution. Together with these core species, we found *Malassezia globosa* in 93.65% of samples and confirmed it by coverage analysis (**Supplementary Fig. 11**). Although *M. globosa* is a known colonizer of the skin, its metagenomic characterization highlights the ability of MetaPhlAn2 to identify non-prokaryotic species. Phages (e.g., for *Propionibacterium*) and double-stranded DNA viruses of the *Polyomavirus* genus were also consistently detected. We subsequently profiled the whole set of 982 samples from other body sites from the Human Microbiome Project (HMP), including 219 samples sequenced after the initial publication (**Supplementary Note 6** and **Supplementary Fig. 12**).

Microbes have been tracked across samples extensively with culture-dependent approaches, and MetaPhlAn2 now offers this possibility in a culture-independent setting by fingerprinting the microbiome at the strain level. This is illustrated by the multiple-time-point ($n = 3$) HMP data set, in which we found that species-specific strain fingerprints were subject specific and conserved longitudinally (**Supplementary Note 7** and **Supplementary Figs. 13–21**). This confirms both strong subject-specific strain retention in the gut microbiome and the ability of MetaPhlAn2 to perform strain fingerprinting and tracking, as these retention patterns are unlikely to occur in longitudinal samples by chance. Additionally, strain identification is possible when a sample contains a previously sequenced genome (**Supplementary Note 7**, **Supplementary Table 13** and **Supplementary Fig. 22**). MetaPhlAn2's enhanced taxonomic profiling (including associated post-analysis, conversion and visualization tools) and its characterization of the HMP data set should serve as convenient tools and extended references for future analysis of the human microbiome.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.3589).

ACKNOWLEDGMENTS
We thank S. Farina and M. Cristofolini for skin-sample collection, V. De Sanctis and R. Bertorelli for metagenomic sequencing, B. Ren for help in the generation of the synthetic metagenomes, the Human Microbiome Project Consortium, and the individuals from the Saint Louis, MO, and Houston, TX, areas, whose generous participation made the HMP possible. This work was supported in part by the US National Institutes of Health (grants R01HG005969 and U54DE023798 to C.H.), the US National Science Foundation (grant DBI-1053486 to C.H.), the US Army Research Office (grant W911NF-11-1-0473 to C.H.), Danone Research (grant PLF-5972-GD to G.W.) the European Union Seventh Framework Programme (Marie Curie grant PCIG13-618833 to N.S.), the Italian Ministry of Education, University and Research (grant FIR RBFR13EWWI to N.S.), Fondazione Caritro (grant Rif.Int.2013.0239 to N.S.) and Terme di Comano (grant to N.S.).

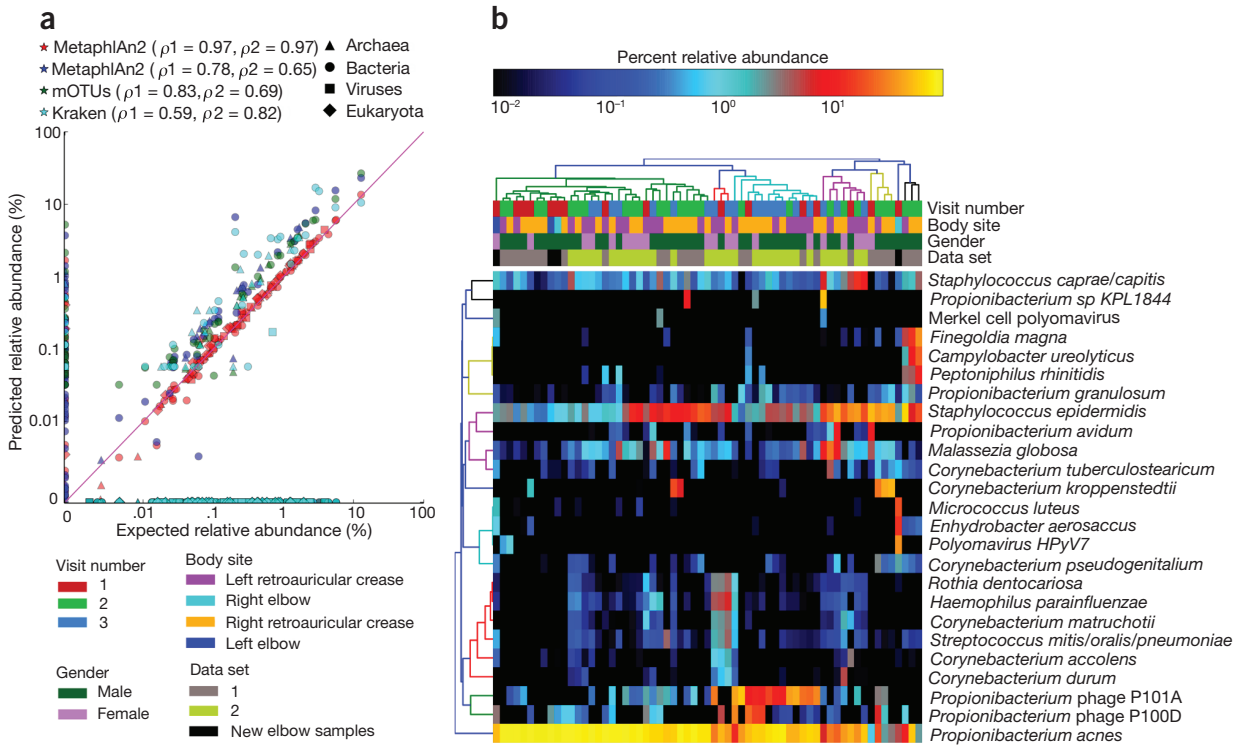

Figure 1 | MetaPhlAn2 can accurately reconstruct the taxonomic composition of shotgun metagenomes. (**a**) When tested on complex (40 million 101-nt-long reads) metagenomes containing bacterial, archaeal, eukaryotic and viral organisms (50, 25, 25 and 25 strains, respectively) as well as 25 strains from unknown species (i.e., without reference genomes in the database), MetaPhlAn2 proved more accurate than MetaPhlAn, mOTU and Kraken (for these three methods, correlations are reported for bacterial and archaeal organisms only, as their support for viral and eukaryotic profiling is limited). (**b**) MetaPhlAn2 characterization of all skin shotgun metagenomes available to date from the HMP and newly sequenced samples.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Duy Tin Truong[1], Eric A Franzosa[2,3], Timothy L Tickle[2,3], Matthias Scholz[1], George Weingart[2], Edoardo Pasolli[1], Adrian Tett[1], Curtis Huttenhower[2,3] & Nicola Segata[1]

[1]Centre for Integrative Biology, University of Trento, Trento, Italy. [2]Biostatistics Department, Harvard School of Public Health, Boston, Massachusetts, USA. [3]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
e-mail: nicola.segata@unitn.it

1. Segata, N. *et al. Nat. Methods* **9**, 811–814 (2012).
2. The Human Microbiome Project Consortium. *Nature* **486**, 207–214 (2012).
3. Scher, J.U. *et al. Elife* **2**, e01202 (2013).
4. Sunagawa, S. *et al. Nat. Methods* **10**, 1196–1199 (2013).
5. Wood, D.E. & Salzberg, S. *Genome Biol.* **15**, R46 (2014).
6. Grice, E.A. *et al. Science* **324**, 1190–1192 (2009).