based methods performed poorly. If closely related genomes were available, the performance of all methods became more similar, with a slight advantage for alignment-based approaches. We observed this for simulated data and the predominant genera of two human gut metagenomes (**Supplementary Tables 5–8**).

PhyloPythiaS also performed well in fragment assignment of 'known unknowns', for organisms of taxonomic clades with no available reference sequence. In this case we observed less 'overbinning', meaning assignments to correct higher-level clades but incorrect low-level clades, than for PhymmBL (**Supplementary Tables 2** and **9–11**). For short fragments of 'known unknowns' (**Supplementary Table 12**), all methods had comparably low assignment accuracy, with MEGAN performing best (**Supplementary Fig. 3** and **Supplementary Table 13**).

Empirical analysis of execution times determined that PhyloPythiaS required 0.08–0.1 seconds for the assignment of 0.1–10 kb fragments (**Fig. 1c**). This was a 3–46-fold and 5–68-fold improvement in comparison to MEGAN and PhymmBL, respectively (**Fig. 1c**). For characterization of a 13-Mb assembled metagenome sample, PhyloPythiaS showed 22-fold, 85-fold and 106-fold speed increase in comparison to PhyloPythia, MEGAN and PhymmBL, respectively (**Supplementary Table 14**). As PhyloPythiaS models require only a subsample of the reference data for accurate assignment, in the future, training times will not necessarily be impacted by increases of sequence data, contrary to alignment-based approaches.

PhyloPythiaS uses an ensemble of linear models whose parameters are identified using the paradigm of support vector machines with structured output spaces to represent composition-based clade specifics of the taxonomic hierarchy instead of an ensemble of multiclass support vector machines for different taxonomic ranks and fragment lengths, as our previously described method PhyloPythia (**Supplementary Note**). PhyloPythiaS has considerable gains in learning and prediction times, and performed similarly to PhyloPythia by several independent measures on two real-world metagenome datasets (**Supplementary Figs. 4** and **5** and **Supplementary Tables 1–3**, **5–7**, **9** and **10**). PhyloPythiaS is freely available for academic use.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank T. Joachims for the SVMstruct implementation, L. Steinbrück and L. Feuerbach for software testing. K.R.P. and A.C.M. were funded by the Max Planck society and Heinrich Heine University Düsseldorf. T.S. acknowledges support from the German Science Foundation (Deutsche Forschungsgemeinschaft. P.J.T. is funded by US National Institutes of Health (P50 GM068763). The unpublished data for the WG-1 project was collected with support provided by the Office of the Chief Executive (OCE) Science team of Commonwealth Scientific and Industrial Research Organization Australia, in the form of an OCE postdoctoral fellowship (P.B.P.) and OCE Science Leader program (M.M.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Kaustubh R Patil[1], Peter Haider[2], Phillip B Pope[3], Peter J Turnbaugh[4], Mark Morrison[3], Tobias Scheffer[2] & Alice C McHardy[1,5]

[1]Max-Planck Research Group for Computational Genomics and Epidemiology, Max Planck Institute for Informatics, Saarbrücken, Germany. [2]University of Potsdam, Department of Computer Science, Potsdam, Germany. [3]Commonwealth Scientific and Industrial Research Organization Livestock Industries, Queensland Bioscience Precinct, St Lucia, Australia. [4]Harvard Faculty of Arts and Sciences Center for Systems Biology, Cambridge Massacusetts, USA. [5]Department of Algorithmic Bioinformatics, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany.
e-mail: alice.mchardy@uni-duesseldorf.de

1. Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. & Hugenholtz, P. *Microbiol. Mol. Biol. Rev.* **72**, 557–578 (2008).
2. McHardy, A.C. & Rigoutsos, I. *Curr. Opin. Microbiol.* **10**, 499–503 (2007).
3. McHardy, A.C., Garcia-Martin, H., Tsirigos, A., Hugenholtz, P. & Rigoutsos, I. *Nat. Methods* **4**, 63–72 (2007).
4. Tsochantaridis, I., Joachims, T., Hofmann, T. & Altun, Y. *J. Mach. Learn. Res.* **6**, 1453-1484 (2005).
5. Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. *Genome Res.* **17**, 377–386 (2007).
6. Brady, A. & Salzberg, S.L. *Nat. Methods* **6**, 673–676 (2009).

Recovery of intact DNA nanostructures after agarose gel–based separation

To the Editor: Molecular self-assembly using DNA as a structural building block has proven an efficient route for construction of nanoscale objects and arrays of ever-increasing complexity[1]. An important catalyst for advancing the field in recent years has been the scaffolded DNA origami strategy, in which a long 'scaffold' strand derived from a viral genome (M13) can be folded with hundreds of short synthetic 'staple' strands into a variety of custom two- and three-dimensional shapes[2,3]. This technology is being used to develop molecular tools for applications in fields such as structural biology[4], single-molecule biophysics and drug delivery. Many of these applications require a homogenous sample of properly folded nanostructures greatly enriched over the misfolded intermediates and large aggregates characteristic of multilayer DNA-origami self-assembly.

Agarose-gel electrophoresis is now the most effective method available for high-resolution separation of well-folded objects on this size scale, but extraction of intact DNA nanostructures with high yield from the agarose matrix is problematic. Existing methods rely on thermal, chemical and/or mechanical destruction of the agarose gel or else electroelution of the DNA to a solid support, leading to problems of low yield, damage to structures and/or contamination with residual agarose. We modified a DNA electroelution method for recovery of DNA from a standard horizontal agarose-gel electrophoresis apparatus to optimize it for efficient, high-resolution and scalable recovery of large and complex intact DNA nanostructures[5,6]. Our initial attempts to purify DNA nanostructures by electroelution revealed the need for a well-sealed elution bed to eliminate high-conductivity buffer paths that served as escape routes for the nanostructures. To address this problem, we poured a 1–2% agarose resolving gel on top of a thinner and more rigid basement layer of 4% agarose previously set in the gel-casting tray (**Supplementary Fig. 1** and **Supplementary Methods**). Once the sample was sufficiently resolved on our dual-layer agarose system, we cut an elution well in the resolving gel directly in front of the band of interest and filled it with a viscous solution of 30–50% sucrose. The elution well is simple to cut down to the interface with the 4% agarose layer because of the difference in rigidity of the layers, and the seal between the layers adjacent to the elution well is not disturbed. To eliminate high-conductivity paths in buffer above the gel, we maintained the running buffer level even with, or below, the surface of the resolving gel. We eluted the band by electrophoresis of the sample into the sucrose bed where movement of the DNA was slowed enough to allow efficient recovery by UV-light detection and micropipetting.

The identity of the elution buffer has profound consequences for the efficacy of purification. Using a 400-nm-long 6-helix bundle nanostructure as a model to assess purification performance (**Fig. 1a** and **Supplementary Table 1**), we screened three solutes at varying concentrations. Use of glycerol or polyethylene glycol resulted in retarded migration of the DNA band and a slow elution time of 1–3 hours, with inconsistent recovery yields between 20% and 60% (**Supplementary Fig. 2**). We obtained the greatest yields with solutions of 30–50% sucrose (**Supplementary Fig. 3**). ImageJ analysis of the gels for purified 6-helix bundles indicated 71 ± 3% of the well-folded structure could be recovered from the agarose matrix versus 15 ± 5% by the pellet-pestle homogenization method[7]. Our analysis by negative-stain transmission electron microscopy (TEM) also indicated a strong enrichment of the properly folded structures.

To evaluate the compatibility of this purification method using other three-dimensional nanostructures, we folded and purified three objects that reflect the range of complexity and the fragility of more elaborate shapes as well as high heterogeneity of unpurified samples. One shape was a 12-helix bundle (**Fig. 1b**, **Supplementary Fig. 4** and **Supplementary Table 1**), whose folding yields more aggregates than for a 6-helix bundle. Agarose-gel analysis of the purified 12-helix bundles indicated that it was not possible to resolve the well-folded structure from aggregates via ion-exchange chromatography (**Supplementary Fig. 5**), but this separation was successful using agarose-gel–based separation. Another shape was a 6-helix bundle bent into a circle, which we designed using the method of targeted deletions and insertions[8] (**Fig. 1c**, **Supplementary Fig. 6** and **Supplementary Table 1**). The final object was a tensegrity structure[9] (**Fig. 1d** and **Supplementary Table 1**). Purification and analysis of each structure by TEM and agarose-gel electrophoresis indicated enrichment of the properly folded structures and yields of 70%, 50% and 45% for the 12-helix bundle, 6-helix ring and tensegrity structure, respectively—values up to fourfold greater than achieved using the pellet-pestle homogenization method[7].

The use of 800-nm 6-helix bundle heterodimers as an alignment medium for membrane-protein nuclear magnetic resonance experiments[7] requires a relatively high degree of purity and nanotube integrity to achieve a liquid crystalline state. When purified via our agarose gel–based method (**Supplementary Fig. 7**), the 6-helix bundles not only dimerized appropriately but also formed high-quality liquid crystals (assayed using bi-refringence), indicating that the structures retained a high degree of structural integrity.

A continued challenge in the field is the hierarchical construction of larger objects from individual nanostructure building blocks. Because individual components often fold with misfolded intermediates in the mixture, the probability of assembling a multimer free from defects becomes very low without prior purification of the components. Using a 12-helix bundle designed to assemble into a tetramer, we found that if the individual components of a larger oligomerized structure are purified before super-assembly, then that super-assembly can proceed with minimal production of large aggregates (**Supplementary Fig. 8**). Previously reported methods of purification are incompatible with more fragile structures that span larger areas or volumes. For example, recovery and TEM detection of a double-cross tensegrity structure has been achieved only by application of our method (T. Liedl, unpublished data). In a few cases we found that the structural integrity of DNA nanostructures

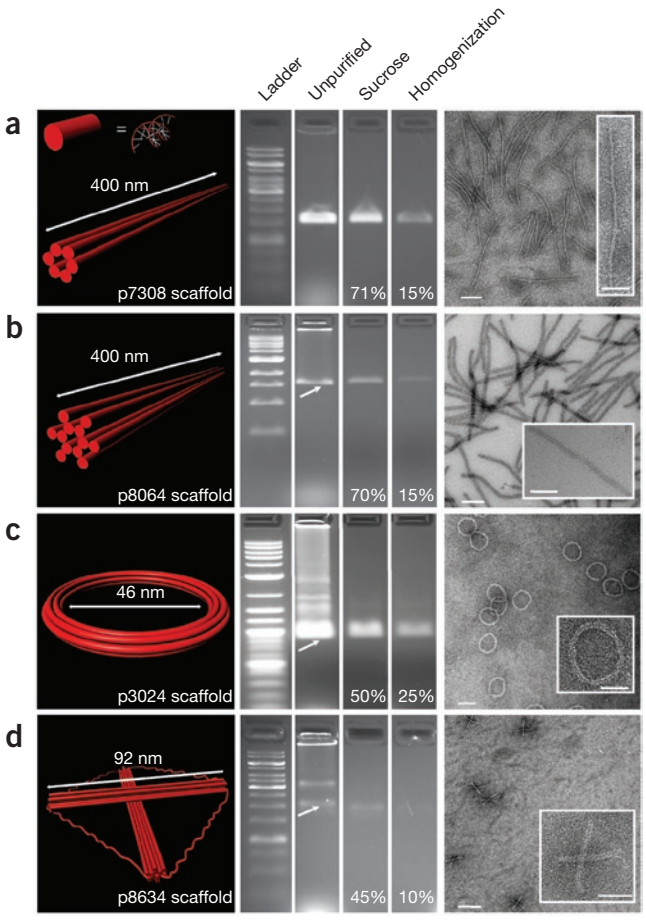

Figure 1 | Agarose-gel and TEM analyses of various DNA origami objects after gel purification. (**a**–**d**) Cylinder models (left; each cylinder represents a DNA double helix) of a 6-helix bundle (**a**), 12-helix bundle (**b**), 6-helix bundle ring (**c**) and prestressed tensegrity-structure kite (**d**). In gel images (middle), lanes for each object were cropped from a single gel. Ladder, kilobase ladder. For unpurified DNA nanostructures (unpurified), arrows indicate the region of each lane that was extracted from the gel during purification before TEM imaging. Also shown are 30% sucrose gel–purified nanostructures (sucrose) and pellet-pestle homogenization–recovered gel-purified nanostructures (homogenization), with estimates of yields after purification indicated. TEM micrographs (right) of the nanostructures after 30% sucrose gel purification. Scale bars, 100 nm (**a**,**b**,**d**), 50 nm (**c**) and in insets, 70 nm (**a**), 80 nm (**b**), 25 nm (**c**) and 50 nm (**d**).

was better preserved when we extracted them using our electrophoresis method instead of the pellet-pestle homogenization method[7] (**Supplementary Fig. 9**). With the method presented here for purifying and oligomerizing larger structures, it should be possible to create more sophisticated three-dimensional DNA nanostructures and DNA liquid crystals.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank S. Simmel for providing TEM micrographs of the double-cross tensegrity structure. This work was funded by grants from the Wyss Institute for Biologically Inspired Engineering, US National Institutes of Health (1DP2OD004641-01 and 1U54GM094608-01) and Office of Naval Research (N000140911118 and N000141010241).

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Gaëtan Bellot[1,2,4], Mark A McClintock[1,2,4], Chenxiang Lin[1–3] & William M Shih[1–3]

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [3]Wyss Institute for Biologically Inspired Engineering at Harvard, Cambridge, Massachusetts, USA. [4]These authors contributed equally to this work.
e-mail: william_shih@dfci.harvard.edu

1. Seeman, N.C. *Annu. Rev. Biochem.* **79**, 65–87 (2010).
2. Rothemund, P.W. *Nature* **440**, 297–302 (2006).
3. Douglas, S.M. *et al. Nature* **459**, 414–418 (2009).
4. Douglas, S.M. Chou, J.J. & Shih, W.M. *Proc. Natl. Acad. Sci. USA* **104**, 6644–6448 (2007).
5. Hansen, H., Lemke, H. & Bodner, U. *BioTechniques* **14**, 28–30 (1993).
6. Maniatis, T., Fritsch, E.F., & Sambrook, J. *A Laboratory Manual* 167 (Cold Spring Harbor Laboratory, 1982).
7. Kurien, B.T., Kaufman, K.M., Harley, J.B. & Scofield, R.H. *Anal. Biochem.* **296**, 162–166 (2001).
8. Dietz, H., Douglas, S.M. & Shih, W.M. *Science* **325**, 725–730 (2009).
9. Liedl, T., Högberg, B., Tytell, J., Ingber, D.E. & Shih, W.M. *Nat. Nanotechnol.* **5**, 520–524 (2010).