Searching for Molecular Solutions – **Cited Notes**

CHAPTER 2

This File contains details on all references to this ftp site within **Chapter 2** of *Searching for Molecular Solutions*. The page numbers of the book where the reference is made are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.

Contents:

Book Reference Page Number	Page Number in this File	Section	
		No.	**Title**
27	2	1	Polymerase activity and exonucleases
58	6	2	Metagenomics

Section 1: *Polymerase activity and exonucleases*

This section provides some more details on the nature of DNA polymerases and their associated exonucleolytic activities. The prototype polymerase for the study of DNA replication is DNA polymerase I of *E. coli* [1]. This enzyme of 928 amino acid residues catalyzes the DNA template-dependent polymerization in a 5' → 3' direction, using deoxynucleotide triphosphates and Mg^{+2} as cofactors. In addition, the enzyme bears two quite distinguishable domains which confer both 5' → 3' and 3' → 5' exonuclease activities. The is manifested as a DNA repair activity, when the enzyme binds at a nick in duplex DNA (Fig. 2Na A).
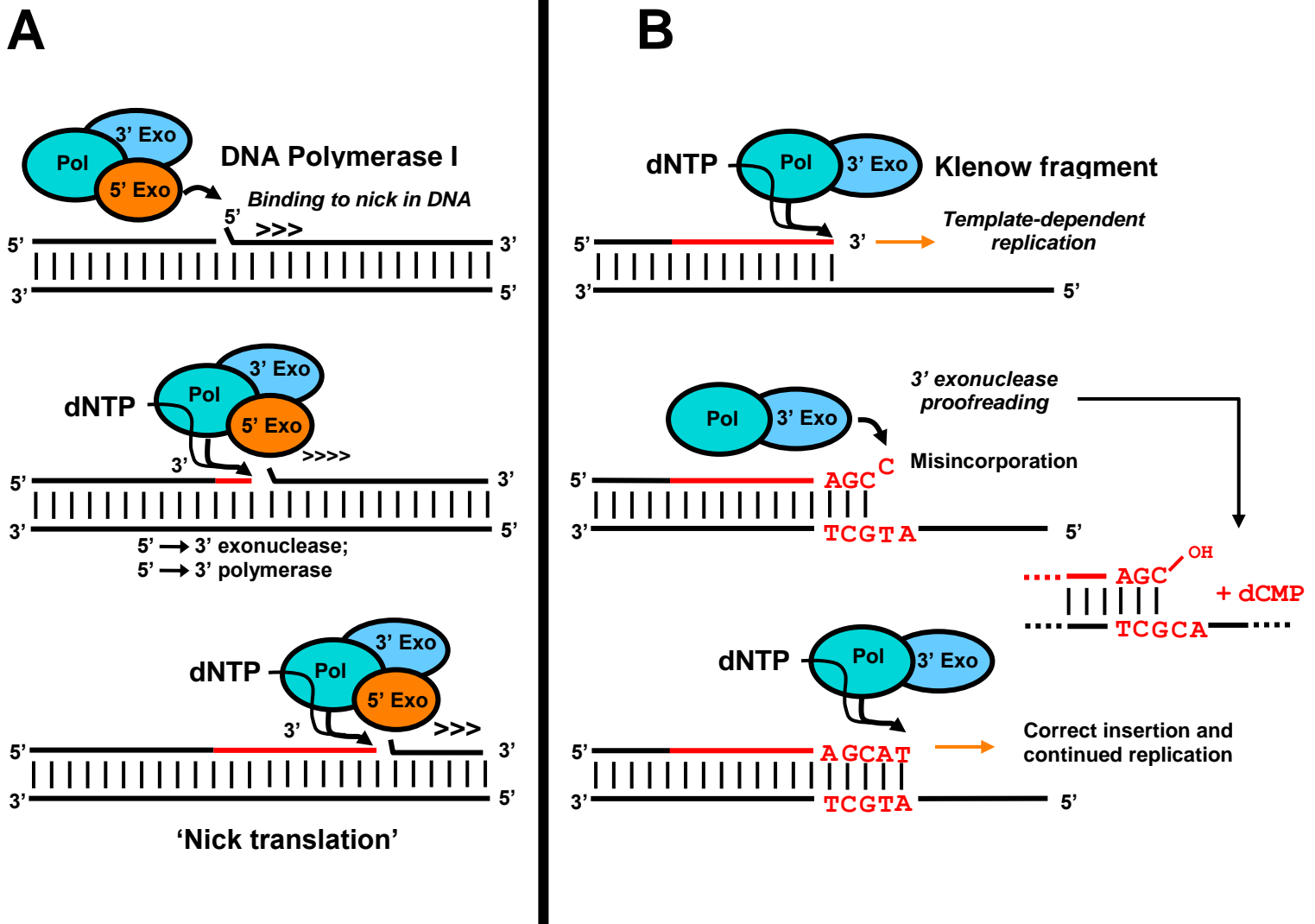

A

DNA Polymerase I

Binding to nick in DNA

5' → 3' exonuclease; 5' → 3' polymerase

'Nick translation'

B

dNTP — **Klenow fragment**

Template-dependent replication

3' exonuclease proofreading

Misincorporation

+ dCMP

Correct insertion and continued replication

Fig. 2.Na

Polymerase and exonucleolytic activities of DNA polymerase I of *E. coli*. Each of these activities is mediated by a separate protein domain, as shown with different colored ovals. **A**, Depiction of the activity of the 5'→ 3' exonuclease activity in conjunction with polymerization. A nick (single-stranded cleavage event) within a DNA duplex is recognized by DNA Polymerase I, and the 5'→ 3' exonuclease activity (active on duplexes in the specified direction) begins to digest the top strand. In conjunction with this, the polymerase function of the enzyme (with deoxynucleotide triphosphates [dNTPs]) initiates extension of the exposed 3' end with newly-synthesized DNA (shown as the top red strand). In this manner, the nick is progressively displaced along the strand, and this 'nick translation' process has long been used as a method for radioactively labeling DNA, by using a duplex partially nicked with DNase I, radiolabeled dNTPs, and the enzyme. B, The activity of the 3'→ 5' proofreading exonuclease activity as contained within the Klenow fragment of DNA Polymerase I (which lacks the 5' → 3' activity). Here a mismatch resulting during polymerization is subject to the 3' → 5' exonuclease, which allows correct base-pairing and DNA synthesis to resume. Note that (in common with all other known biological polymerases) each new template-directed base addition is at the 3' end of the growing newly synthesized chain; accordingly the polymerization direction is 5' → 3'.

The modular domain nature of DNA polymerase I has long been known from the ability to separate the polymerase and 3' → 5' exonucleolytic activities as discrete fragments [2,3]. The latter exonuclease functions as a molecular 'proof-reader', since mispaired bases resulting from misincorporation events are recognized as substrates for 3' → 5' exonucleolytic action, allowing correction of the fault and resumption of polymerization (Fig. 2Na B).

The distinctions between the polymerase and 3' → 5' exonuclease sites for the Klenow fragment are shown structurally in Fig. 2Nb.
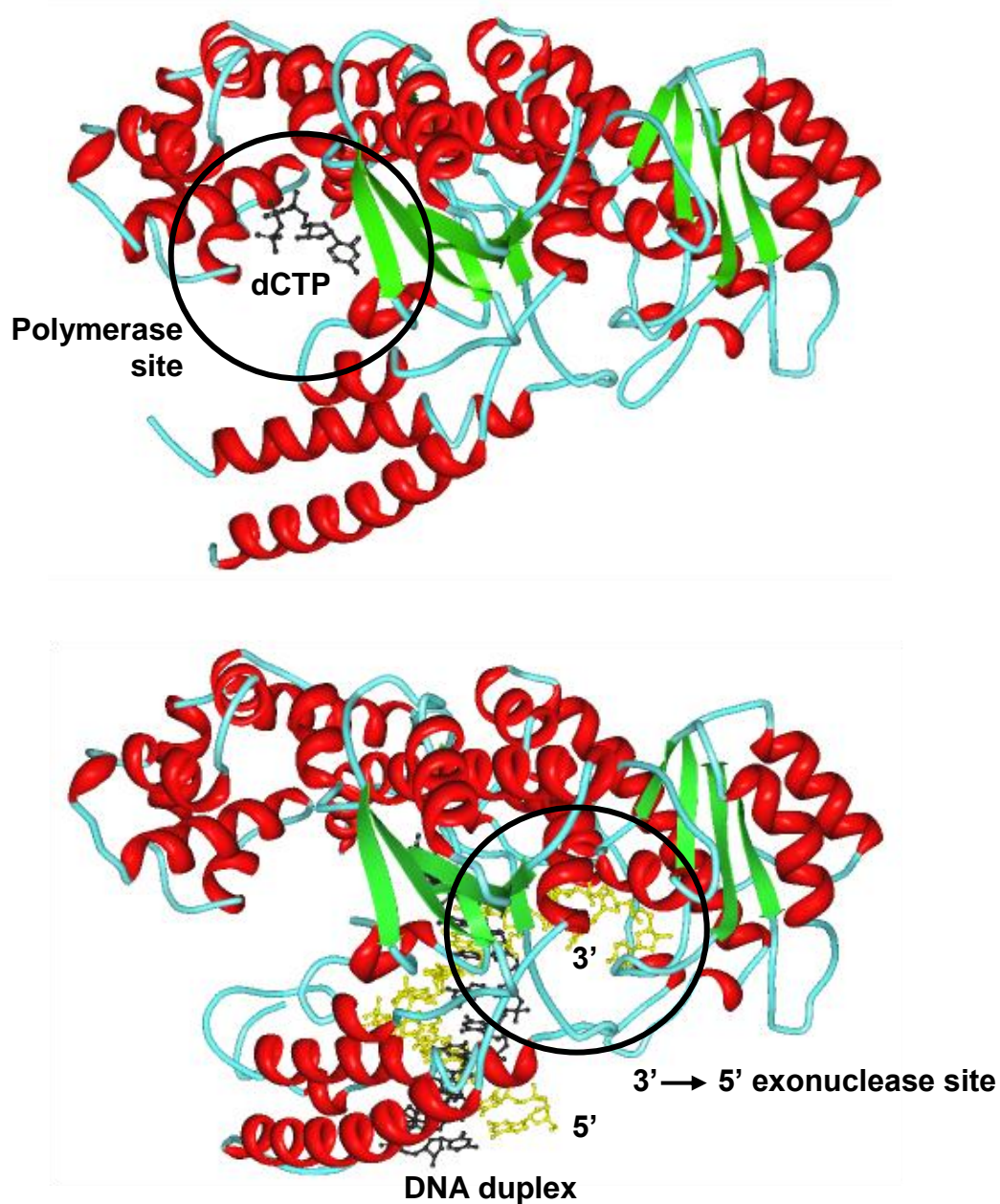
Fig. 2.Nb

Structures of DNA Polymerase I Klenow fragment, showing distinct domain sites for polymerase and 3' $\rightarrow$ 5' exonucleolytic activities. In both cases, α-helices red, β-strands greens, loops and turns light blue. TOP, X-ray structure with co-crystallized dNTP (here, specially dCTP) in the cleft of the polymerase domain [4]; BOTTOM, complex of Klenow fragment with a DNA duplex with a 3-base 3' extension; held within a groove in the 3' $\rightarrow$

5' exonuclease domain. [5] (DNA strand with the 3' extension yellow; complementary strand black). In this case, the enzyme has a mutation (D355A) which greatly reduces the actual exonucleolytic activity [6,7], in order to avoid hydrolysis of the strand during crystallization [5]. Sources: Protein Data Bank [8] 1KFD (Top) and 1KDS (Bottom). Images generated with Protein Workshop [9].

A DNA primer-template can shuttle between replication within the polymerase domain and editing in the Klenow fragment exonuclease domain. This can occur in either an intermolecular fashion (that is, with dissociation from an enzyme molecule and re-binding by another) or intramolecular (shuttling from one site to another in the same enzyme molecule) [5,10]. Fig. 2Na can be interpreted along the lines of either inter- or intramolecular activities.

Section 2:   ***Metagenomics***

Although micro-organisms have contributed an enormous number of useful bioproducts, it has been estimated that only ~1% of all microbial species are currently known [11]. Bacteria can thrive in surprisingly diverse and extreme environments, and characterization of full microbial populations and ecologies under such conditions may not be a simple exercise. But in order to study pure populations of microbes in the laboratory, it is necessary to propagate them. From the soil and other complex environments, as low as 0.2% of the total set of microbial species may be cultivatable [12,13]. Recent advances have shown that many 'uncultivatable' microbes may indeed be grown in the laboratory under the right conditions [14,15], but systematic propagation of environmental organisms remains a major constraint for the field as a whole. One way around this limitation is to 'shotgun' clone ♥ pooled DNAs from soil (or other) microbial habitats into a tractable *E. coli* bacterial strain, and screen for the production of novel bioproducts. DNA samples taken directly from environmental specimens can consequently be used for genomic studies of non-cultivatable organisms, a powerful approach which has come to be termed 'metagenomics'.

Applications of the metagenomic strategy are increasing rapidly [16], and its potential for the acceleration of natural molecular discovery has been highlighted [17,18] along with the accompanying expansion of known biological sequence space [19] and ecological knowledge [20,21]. Any complex environmental microbial population is in principle amenable to this kind of analysis, including soil [22], physiological sites such as the human gut [23,24], extreme environments (as frequented by the 'extremophile' organisms referred to above [25]), and the

---

♥ The term 'shotgun cloning' has been around since the early days of molecular biology, and refers to the random joining of fragments of an organism's DNA into a replicable vector which allows the fragments to be grown at will in a bacterial host.

oceans[26]. Some important upshots of such undertakings have been an extension of the known number of protein families [26], and greatly improved knowledge of the phosphorylation-based signaling pathways used by prokaryotes [27]. Metagenomics with a 'bulk' environmental DNA sample will include all micro-organisms present, and their viral parasites. However, size-based (filtration) and other techniques can be used to enrich for viral nucleic acids, such that informative studies of bacteriophage environmental 'metaviromes' have also been performed [28-30]. But with modern high-throughput sequencing, a 'shotgun' approach to metagenomics has the potential to distinguish novel adventitious parasites from a background of host genomic sequences. A notable example of this approach is the identification of a candidate viral pathogen for the colony collapse disorder of honeybees [31].

By its nature, a comprehensive metagenomic DNA library will carry a diversity of biological information, and two major ways of extracting this informative content have been used: sequence analysis and function-driven screening [32,33]. Bioinformatic studies, as noted with the above marine metagenomics, need sequence information as their raw material and fall into the first of these categories. Screening for a specific sought-after function also requires appropriate sequences, but is more challenging in that the required sequence information must be accurately expressed, preferably in a manner allowing high-throughput screening of very large numbers of clones. In the case of complex metabolite syntheses, the challenge is further upgraded by the need to simultaneously clone and express multiple genes in the biosynthetic pathway. A powerful approach which is often applicable in this area is to use bacterial hosts (usually, but not always, *E. coli* ♥) which lack (or have mutagenized) the pathway of interest, and then to search for hosts bearing relevant cloned metagenomic

---

♥ Such has been the ubiquity of *E. coli* as a molecular biological work-horse, that it has been said in the past that if one were to pursue molecular biology, one was necessarily interested in at least two cells: *E. coli* and the other system worked on. With *in vitro* cloning methods and alternative hosts, this dictum has less impact now, but is still widely applicable.

DNA with an appropriate selective medium [33]. In such cases, the premise is that a few clones of environmental DNA samples (out of a vast number in total, constructed in plasmid replicating vehicles) will allow the expression of the gene product (nutrient) needed for growth of the mutant bacterial host. Host cells which have taken up the appropriate clone and express the otherwise-missing nutrient can then be isolated on a defined growth medium lacking the specific nutrient.

For functional expression of an entire biosynthetic pathway in a foreign bacterial host to be successful, three basic requirements must be fulfilled: (1) all of the gene sequences for the complete pathway must be present, including regulatory sequences directing their expression; (2) the latter control sequences must be functional in the bacterial host, and (3) if the bioproduct is self-toxic in its original prokaryotic context in the absence of specific self-modification, another gene product conferring protection to the host organism must also be transferred to the new bacterial host. For the first requirement to be satisfied, it is generally necessary to ensure that large contiguous segments from the environmental DNA pool are used [34]. The second condition is often the case (at least if the donor DNA itself is bacterial in origin), and the third is a logical necessity, as noted in Chapter 3 of *Searching for Molecular Solutions* (p. 64).

One current limitation of some metagenomic ambitions is in fact isolation of environmental DNAs, whose quality and size can be compromised by co-purifying contaminants [13]. Where expression of biosynthetic proteins from metagenomic sources may be problematic for the traditional vehicle of *E. coli*, other microbial host such as species of *Streptomyces* [13] or *Pseudomonas* [35] may be chosen as alternatives. Even better, perhaps, is the option of avoiding using a bacterial host altogether. This has been possible in principle since the advent of the polymerase chain reaction (PCR; See SMS-CitedNotes-Ch4 / Section 6; from the same ftp site), and can be performed in the absence of sequence information for the target DNAs by means of ligating on specific linker-primer segments.

Single molecules can be cloned by PCR amplification, which is achievable by *in situ* methods resulting in 'PCR colonies' or 'polonies' [36,37], also known as 'molecular colonies' [38,39]. Despite these and other technical advances which have extended the range of PCR, this technique is still limited in the size of the target sequences which can be effectively amplified, especially if performed from single-molecule templates (as required for cloning purposes). A potential alternative to PCR for metagenomic applications is the use of isothermal amplifications with a bacteriophage DNA polymerase (phage ø29). The properties of this polymerase are such that under the right conditions it can accurately replicate circular DNA templates by 'rolling circle amplification' and allow '*in vitro* cloning' [40], albeit with certain remaining technical limitations [41].

Nonetheless, *in vitro* cloning technologies put a different spin onto metagenomic studies. Instead of sorting through 'bulk' cloned libraries of DNAs from environmental sources, it will become increasingly feasible to alternatively isolate single cells (bacterial or other) from the same environmental origin. With an effective means for *in vitro* cloning of such a cell's genome [41,42], the need for propagation of the organism becomes superfluous. Indeed, *in vitro* cloning with polymerases ('ploning') permits full-genome sequencing and accompanying genomic studies [43]; a new era in 'single-cell genomics' [44].

References

1.   Kornberg, A. & Baker, T. S. *DNA Replication. 2nd Edition.* (University Science Books, 2005).

2.   Klenow, H. & Henningsen, I. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. *Proc Natl Acad Sci U S A* **65**, 168-75 (1970).

3.   Klenow, H. & Overgaard-Hansen, K. Proteolytic cleavage of DNA polymerase from Escherichia Coli B into an exonuclease unit and a polymerase unit. *FEBS Lett* **6**, 25-27 (1970).

4.   Beese, L. S., Friedman, J. M. & Steitz, T. A. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32**, 14095-101 (1993).

5.   Beese, L. S., Derbyshire, V. & Steitz, T. A. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**, 352-5 (1993).

6.   Beese, L. S. & Steitz, T. A. Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo J* **10**, 25-33 (1991).

7.   Derbyshire, V., Grindley, N. D. & Joyce, C. M. The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. *Embo J* **10**, 17-24 (1991).

8.   Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980 (2003).

9.   Moreland, J. L., Gramada, A., Buzko, O. V., Zhang, Q. & Bourne, P. E. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**, 21 (2005).

10.   Joyce, C. M. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J Biol Chem* **264**, 10858-66 (1989).

11.   Cragg, G. M. & Newman, D. J. Medicinals for the millennia: the historical record. *Ann N Y Acad Sci* **953**, 3-25 (2001).

12.   Pettit, R. K. Soil DNA libraries for anticancer drug discovery. *Cancer Chemother Pharmacol* **54**, 1-6 (2004).

13.   Streit, W. R., Daniel, R. & Jaeger, K. E. Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Curr Opin Biotechnol* **15**, 285-90 (2004).

14.   Zengler, K. et al. Cultivating the uncultured. *Proc Natl Acad Sci U S A* **99**, 15681-6 (2002).

15.   Podar, M. et al. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl Environ Microbiol* **73**, 3205-14 (2007).

16.   Ward, N. New directions and interactions in metagenomics research. *FEMS Microbiol Ecol* **55**, 331-8 (2006).

17.   Langer, M. et al. Metagenomics: an inexhaustible access to nature's diversity. *Biotechnol J* **1**, 815-21 (2006).

18.   Schmeisser, C., Steele, H. & Streit, W. R. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* **75**, 955-62 (2007).

19.   Cowan, D. et al. Metagenomic gene discovery: past, present and future. *Trends Biotechnol* **23**, 321-9 (2005).

20.   Tringe, S. G. et al. Comparative metagenomics of microbial communities. *Science* **308**, 554-7 (2005).

21.   DeLong, E. F. Microbial community genomics in the ocean. *Nat Rev Microbiol* **3**, 459-69 (2005).

22.   Daniel, R. The soil metagenome--a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* **15**, 199-204 (2004).

23.   Gill, S. R. et al. Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-9 (2006).

24.   Zoetendal, E. G., Vaughan, E. E. & de Vos, W. M. A microbial world within us. *Mol Microbiol* **59**, 1639-50 (2006).

25.   Ferrer, M., Golyshina, O., Beloqui, A. & Golyshin, P. N. Mining enzymes from extreme environments. *Curr Opin Microbiol* **10**, 207-14 (2007).

26.   Yooseph, S. et al. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. *PLoS Biol* **5**, e16 (2007).

27.   Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. & Manning, G. Structural and Functional Diversity of the Microbial Kinome. *PLoS Biol* **5**, e17 (2007).

28.   Paul, J. H. & Sullivan, M. B. Marine phage genomics: what have we learned? *Curr Opin Biotechnol* **16**, 299-307 (2005).

29.   Angly, F. E. et al. The marine viromes of four oceanic regions. *PLoS Biol* **4**, e368 (2006).

30.   Casas, V. & Rohwer, F. Phage metagenomics. *Methods Enzymol* **421**, 259-68 (2007).

31.   Cox-Foster, D. L. et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283-7 (2007).

32.   Schloss, P. D. & Handelsman, J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14**, 303-10 (2003).

33.   Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**, 669-85 (2004).

34. Bertrand, H. et al. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J Microbiol Methods* **62**, 1-11 (2005).

35. Ono, A. et al. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl Microbiol Biotechnol* **74**, 501-10 (2007).

36. Mitra, R. D. & Church, G. M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27**, e34 (1999).

37. Mitra, R. D. et al. Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci U S A* **100**, 5926-31 (2003).

38. Samatov, T. R., Chetverina, H. V. & Chetverin, A. B. Expressible molecular colonies. *Nucleic Acids Res* **33**, e145 (2005).

39. Chetverin, A. B. & Chetverina, H. V. Molecular colony technique: a new tool for biomedical research and clinical practice. *Prog Nucleic Acid Res Mol Biol* **82**, 219-55 (2008).

40. Hutchison, C. A., 3rd, Smith, H. O., Pfannkoch, C. & Venter, J. C. Cell-free cloning using phi29 DNA polymerase. *Proc Natl Acad Sci U S A* **102**, 17332-6 (2005).

41. Stepanauskas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* **104**, 9052-7 (2007).

42. Zhang, K. et al. Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24**, 680-6 (2006).

43. Zhang, K. et al. Long-range polony haplotyping of individual human chromosome molecules. *Nat Genet* **38**, 382-7 (2006).

44. Hutchison, C. A., 3rd & Venter, J. C. Single-cell genomics. *Nat Biotechnol* **24**, 657-8 (2006).