Journal of Engineering and Technology Research Vol. 3(8), pp. 284-286, 12 September, 2011
Available online at http:// www.academicjournals.org/JETR
ISSN 2006-9790 ©2011 Academic Journals

Short Communication

Modeling the distribution of primes in computable biomolecules

Okunoye Babatunde O

Department of Pure and Applied Biology, Ladoke Akintola University of Technology, Ogbomoso, Nigeria. E-mail: babatundeokunoye@yahoo.co.uk. Tel: + 234-08036296145.

Accepted 28 July, 2011

Prime numbers are an important research area with application in computer security. The security of certain cryptographic schemes critical in information security lies in the difficulty of factoring large numbers into their prime factors. Determination of the distribution of prime numbers is an important unsolved problem in number theory. Biomolecules such as DNA have extended the domains of what is considered a standard computer model. This report investigated the distribution of prime numbers generated from viral DNA. The digits of Euler's number and Pi were found to be encoded between Bacteriophage T4 Watson-Crick DNA segments comprised of prime numbers.

Key words: Prime number, number theory, molecular computing, phage T4.

INTRODUCTION

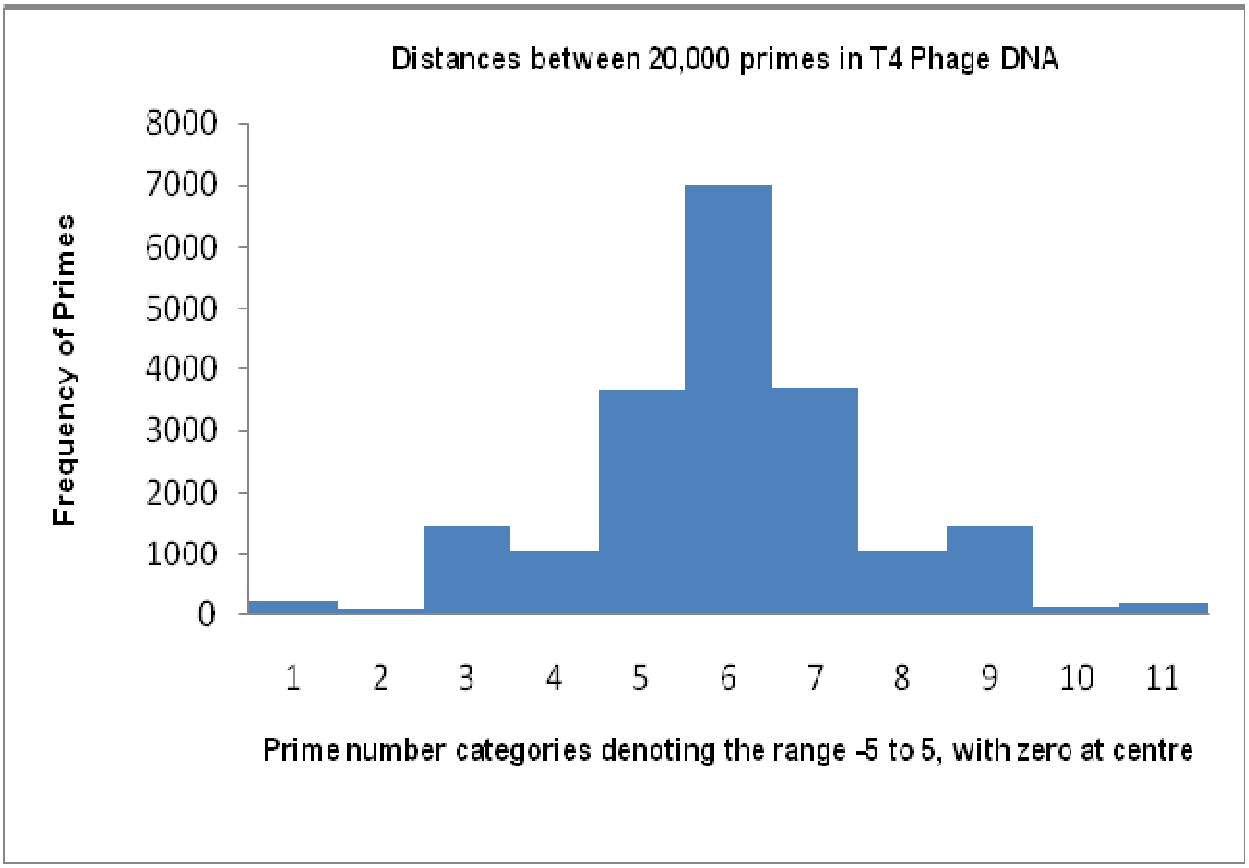
A prime number is a natural number larger than one which cannot be expressed as a product of two smaller natural numbers (Tao, 2007). The prime numbers less than 20 are 2, 3, 5, 7, 11, 13, 17, and 19. Prime numbers, once thought to be of interest only to number theorists, have found application in computer science and technology, mainly in the protocols of information security and cryptography – with the development of public key cryptography (Rivest et al., 1978), whose security lies in the difficulty of factoring large numbers into their prime factors. The most common public key encryption algorithm is RSA, named after its developers Ron Rivest, Adi Shamir and Leonard Adleman (Rivest et al., 1978).

The standard computer model in computability theory is the Turing Machine (Turing, 1936). Deoxyribonucleic acid (DNA) has been used in the implementation of the Turing Machine (Rothemund, 1996; Qian et al., 2011), birthing the field of molecular computing (Lipton, 1995; Winfree, 1996; Quyang et al., 1997; Lagoudakis and LaBean, 2000; Braich et al., 2002; Rothemund et al., 2004). Molecular computing has brought biological molecules into the field of computer science and engineering. DNA computation also has potential application in fabrication tasks in nanotechnology (Cook et al., 2004; Barish et al., 2005) and in the design of digital logic circuits (Seelig et al., 2006).

Determining the distribution of prime numbers among natural numbers is an important unsolved problem in mathematical number theory (Tao, 2007). The question is closely linked with the Riemann Hypothesis (Conrey, 2003), which predicts a precise formula for the probability distribution of primes. The distribution of 20,000 prime numbers generated from Bacteriophage T4 DNA was investigated, with the objective of determining their distribution. Euler's number (*e*) and Pi (π) are two of the most important mathematical constants and a part of Euler's identity: $e^{i\pi} + 1=0$. The value of *e* and π to ten significant figures respectively are e=2.718281828 and π=3.141592653, obtained from the Online Encyclopedia of Integer Sequence (Sequences A0011133 and A000796 respectively).

MATERIALS AND METHODS

Bacteriophage T4 Genome (Miller et al., 2003) was sourced from GenBank, the institutional genome depository with accession number AF158101. Bacteriophage T4 represents the most understood model for modern genomics and proteomics; and its study has revealed many insights and paradigms in molecular biology. The numbers of each nucleotide base per Watson-Crick single strand DNA segments, comprising of ten bases each were counted and recorded in successive single strand segments of T4 phage DNA (single strand complement 1' – 99,180' in the 3' to 5' direction). The prime numbers present as numbers of nucleotides were: 2, 3, 5 and 7; and the distances between them in 20,000 numbers generated from T4 Phage DNA was investigated. The frequency histogram of the distances was plotted with Microsoft® Excel statistical package.T4 Phage Watson-Crick single strand DNA segments comprising only of prime numbers consist of the



Distances between primes (x)	-5	-4	-3	-2	-1	0	1	2	3	4	5
Frequency (f)	215	90	1455	1051	3662	7018	3696	1059	1446	119	189

Figure 1. Histogram showing the frequency distribution of the distances between prime numbers generated from T4 Phage DNA Watson-Crick units.

following combinations of nucleotide bases: 55, 37, 235 and 2233. The spatial distances between these prime segments were recorded and these distances traced out digits of the mathematical constants pi and e.

RESULTS

The distribution of the 20,000 prime numbers generated from Bacteriophage T4 nucleotide bases followed a binomial distribution (Figure 1). The mean and standard deviation of a binomial random variable is expressed as: $\mu = np$ and $\sqrt{np(1-p)}$. An understanding of the distribution of prime numbers among the natural numbers – a subject of the Riemann Hypothesis, is listed by the Clay Mathematics Institute as one the seven most important open questions in Mathematics and a Millennium Prize Problem (Sarnak, 2004). This work describes the distribution of prime numbers generated from a biomolecule, DNA. The secondary result of this investigation is that the distances between T4 Phage Watson-Crick DNA segments comprised of prime numbers encoded 12 values of e and 15 values of π to ten significant figures (Table 1).

DISCUSSION AND CONCLUSION

The principal finding of this work is that the distribution of prime numbers in DNA follows a binomial distribution. The study of the nature of prime numbers is of importance in computer security (Tao, 2007). Secondarily, the efforts directed at computing the values of mathematical constants such as pi and e (Shanks and Wrench, 1962; Wei et al., 1996; Bailey et al., 1997) permits us two ways of making sense of the encoding of the two mathematical constants within DNA molecules. On the one hand, there is merit in attributing this to pure chance; on the other hand, it might point to non-trivial behaviour in DNA molecules, which is not a surprise given the new role of DNA molecules in computing.

Table 1. Spatial distances of digits of e and π as encoded between the distances of Watson-Crick units of T4 DNA comprised of prime numbers.

Spatial distances of digits of e

e	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}
2	26	170	270	410	620	702	780	896	1014	1173	1401	1571
7	35	175	280	428	630	714	789	915	1032	1174	1419	1575
1	36	183	281	431	633	723	814	921	1039	1178	1420	1579
8	69	196	293	571	635	731	865	929	1047	1191	1427	1589
2	90	202	300	572	650	735	866	931	1049	1204	1466	1594
8	103	222	308	585	661	750	867	955	1080	1205	1479	1621
1	105	225	320	593	662	753	879	959	1082	1208	1484	1623
8	122	258	348	595	686	758	881	995	1135	1338	1522	1635
2	130	261	351	613	687	760	885	1003	1136	1354	1526	1636
8	168	265	378	619	700	776	890	1007	1161	1398	1559	1642

Spatial distances of digits of π

π	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}
3	23	161	238	429	525	665	786	982	1072	1146	1207	1358	1443	1498	1596
1	27	163	240	431	539	675	814	984	1078	1153	1208	1365	1448	1519	1601
4	37	179	288	433	545	688	818	987	1086	1155	1216	1369	1449	1530	1619
1	40	183	295	437	548	689	834	1002	1090	1157	1219	1376	1453	1532	1623
5	57	190	297	447	559	694	839	1008	1092	1158	1234	1379	1469	1535	1627
9	109	195	337	455	601	742	883	1040	1102	1177	1258	1411	1473	1561	1683
2	119	202	351	458	613	744	885	1049	1113	1190	1282	1412	1478	1571	1689
6	123	207	385	463	616	763	965	1057	1116	1195	1350	1413	1483	1576	1690
5	135	224	401	512	618	768	976	1061	1137	1196	1352	1437	1490	1580	1691
3	147	227	402	513	654	782	979	1070	1144	1198	1355	1441	1491	1583	1706

REFERENCES

Bailey D, Borweinand P, Plouffe S (1997). On the Rapid Computation of Various Polylogarithmic constants. Math. Comput., 66: 903–913.

Barish RD, Rothemund PWK, Winfree E (2005). Two computational primitives for algorithmic self-assembly: Copying and counting. Nano Lett., 5: 2586-2592.

Braich RS, Chelyapov N, Johnson C, Rothemund PWK, Adleman L (2002). Solution of a 20-variable 3-SAT Problem on a DNA Computer. Science, 296: 499-502.

Conrey JB (2003).The Riemann Hypothesis. Notices Am. Math. Soc., 50: 341-353.

Cook M, Rothemund PWK, Winfree E (2004). Self-assembled circuit patterns. In Chen J and Reif J (eds), DNA Computers, volume 9. Lect. Notes Comput. Sci., 2943, Berlin Heidelberg, Springer-Verlag, pp. 91-107.

Lagoudakis MG, LaBean TH (2000). 2-D DNA self-assembly for satisfiability. In Winfree E and Gifford DK (eds) DNA Based Computers V, volume 54 of DIMACS, Providence, RI. American Mathematical Society, pp. 141-154.

Lipton RJ (1995). DNA solution of hard computational problems. Science, 268: 542-545.

Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W (2003). Bacteriophage T4. Microbiol. Mol. Biol. Rev., 67: 86-156

Qian L, Soloveichik D, Winfree E (2011). Efficient Turing-Universal computation with DNA polymers. (Eds: Sakakibara Y. and Mi Y.) In DNA 16, Lect. Notes Comput. Sci., 6518: 123-140.

Quyang Q, Kaplan PD, Liu S, Libchaber A (1997). DNA solution of maximal clique problem. Science, 278: 446-449.

Rothemund PWK (1996). A DNA and restriction enzyme implementation of Turing Machines. In Lipton RJ and Baum EB (eds) DNA Based Computers, volume 27 of DIMACS, Providence, RI, 1996. American Mathematical Society, pp. 75-120.

Rothemund PWK, Papakakis N, Winfree E (2004). Algorithmic self-assembly of DNA Sierpinski triangles. PLoS Biol., 2: 424-436.

Rivest R, Shamir A, Adleman L (1978). A method for obtaining digital signatures and Public-Key Cryptosystems. Commun. ACM., 21(2): 120-126.

Sarnak P (2004). Problems of the Millennium: The Riemann Hypothesis. Clay Mathematics Institute, p. 1.

Seelig G, Soloveichik D, Zhang DY, Winfree E (2006). Enzyme-free nucleic acid logic circuits. Science, 314: 1585–1588.

Shanks D, Wrench JW (1962). Calculation of Pi to 100,000 decimals. Math. Comput., 16: 76-79.

Tao T (2007). Structure and randomness in the prime numbers: A small collection of results in number theory. UCLA Science Colloquium, pp. 2, 24, 18.

Turing A (1936). On computable numbers with an application to the entscheidnungsproblem. Proc. London Math. Soc., 42(2): 230–265.

Wei GY, Yang ZG, Sun JC, Li JK (1996). The computation of Pi to 10,000,000 decimal digits. J. Numer. Methods Comput. Appl., 17: 78-81.

Winfree E (1996). On the computational power of DNA annealing and ligation. In R. J. Lipton RJ and Baum EB (eds) DNA Based Computers, volume 27 of DIMACS, Providence, RI, 1996. American Mathematical Society, pp. 199-221.