研究报告 Original Papers

四种禾本科作物叶绿体基因组碱基替换的侧翼序列特征

黄卓然, 吴晓敏, 张慧君, 段永波, 张强*

淮北师范大学生命科学学院, 安徽淮北235000

摘要: 对于核苷酸替换对相邻位点依赖性及侧翼序列特征的研究有助于厘清不同物种之间的进化关系, 可以作为基因编辑和基因修饰技术的基础。早期研究表明, CpG甲基化效应广泛存在于哺乳动物和细菌基因组中, 而叶绿体某些基因中的颠换相邻位点表现出一定的碱基组成偏好。本研究将4种禾本科作物小麦(*Triticum aestivum*)、水稻(*Oryza sativa*)、玉米(*Zea mays*)和高粱(*Sorghum bicolor*)的叶绿体基因组序列分为两组比对, 查找替换位点并统计侧翼序列各位点碱基组成。通过相对熵(KL散度)作图, 发现小麦和水稻的叶绿体基因组中相邻序列的碱基组成受到替换的影响随着距离的增大而减小, 而玉米和高粱叶绿体基因组缺少相应规律。对于相对熵较高的替换相邻位点, 通过不同核苷酸相对熵贡献的研究, 发现4种禾本科作物的叶绿体基因组均表现出CpG甲基化效应以及颠换相邻位点的特殊组成规律。本研究为相对熵在植物基因组分子进化领域的应用提供参考。

关键词: 禾本科作物; 叶绿体基因组; 替换位点; 侧翼序列; 相对熵

禾本科植物分布广泛, 具有极高的经济和生态价值。为人们所熟知的禾本科植物包括小麦(*Triticum aestivum*)、水稻(*Oryza sativa*)、玉米(*Zea mays*)、高粱(*Sorghum bicolor*)等都是常见的粮食作物。作为植物细胞内重要的细胞器, 叶绿体一直是禾本科植物研究的重点。根据“内共生起源”假说, 叶绿体起源于被真核细胞所吞噬的光合细菌(张秀月和印莉萍2017), 而叶绿体基因组的结构、大小和组成等确实具有与光合细菌相似的特征(Mader等2018)。相比于植物细胞核基因组而言, 叶绿体基因组虽小但保守性更强, 且完成的功能较为独立和集中, 从分子进化的角度来看是非常好的研究材料(员涛等2015; Dong等2018)。自从1986年烟草(*Nicotiana tabacum*)叶绿体基因组测序完成以来(Shinozaki等1986), 关于叶绿体基因组的研究不断得到推进。随着测序技术的发展, 尤其是高通量测序技术出现之后, 叶绿体基因组数据库不断得到补充和完善。其中, 重要的禾本科作物叶绿体基因组已经基本完成了测序(Tsuruta等2017), 这为本研究提供了重要的数据支持。

核苷酸替换是指物种之间存在的在基因组的某一位点或多个位点所发生的核苷酸改变, 共12种替换类型, 可概括为转换和颠换两大类。物种内个体之间产生的核苷酸突变被遗传保留下来之后, 可能导致后代出现性状改变而分化成不同种群, 核苷酸替换由此产生(Long等2018; Sprouffske等2018)。对于核苷酸替换的相邻位点研究是一项重要课题, 它有助于理解核苷酸替换、DNA蛋白质互作以及DNA修复等的基本发生机制(Poelwijk 2019)。目前关于核苷酸替换对相邻位点依赖性的研究中, 由CpG二联子到TpG的突变最受关注(Panchin等2016; Liu等2016; Kessler等2019)。在大肠杆菌(*Escherichia coli*)某些基因中, CpG二联子的胞嘧啶(cytosine, C)容易被甲基化, 继而脱氨基变为胸腺嘧啶(thymine, T), 这种CpG效应就是典型的相邻位点影响核苷酸的替换的例子(Coulondre等1978)。而对叶绿体基因组的研究揭示了鸟嘌呤(guanine, G)突变为腺嘌呤(adenine, A)或者C突变为T受到上下游序列AT含量的影响规律(Zheng等2007)。替换位点与距离更远位点的关系仍存在一定的争议: 一种观点认为核苷酸替换受到侧翼序

收稿 2019-03-26 **修定** 2019-06-17

资助 安徽省高校自然科学研究项目(KJ2016B011)、安徽省高校优秀青年骨干教师国内访问研修项目(gxgnfx2018084)和安徽省自然科学基金(1608085MC49)。

* 通讯作者(zhangq@foxmail.com)。

列的影响范围仅限于上下游各两个位点(Krawczak等1998), 另一种观点则倾向认为替换位点与上下游2~5 bp的序列都有紧密的联系(Nevarez等2010)。然而, 在分析生物序列中替换模式特征的研究中, 之前的方法多为直接将侧翼序列的碱基组成与全基因组序列碱基组成做简单的差异计算, 难以从4种碱基的整体进行特征规律的研究。因此, 本研究引入相对熵[relative entropy, 又称KL散度(Kullback–Leibler divergence)]的研究方法, 与传统方法相比, 相对熵作图可以更为直观地反映两种频率分布之间的差异。信息论的概念最初由美国数学家、电子工程师和密码学家Claude E. Shannon提出, 原本用以系统化分析广义通信系统中的数据流。相对熵是信息论研究中的重要概念, 许多研究者已经开始使用其来分析遗传信息(Ma等2010; Chen等2017)。对于基因组序列来说, 计算得到的信息熵值越小, 意味着遗传信息越稳定, 即序列越保守。而相对熵可以用来衡量两条不同序列或者同一序列在进化过程中的遗传信息差异情况(刘军和许甫荣2003)。

目前对于禾本科植物叶绿体基因组的研究多集中在基因组测序、比较基因组学和系统发育等领域(Sun等2019; Yang等2019), 而从分子进化角度来分析种间核苷酸替换及与上下游序列相关性的研究相对较少。小麦、水稻、玉米和高粱作为我国最常见、种植面积最为广泛的禾本科粮食作物, 对其叶绿体基因组中核苷酸替换规律进行研究将有助于明确其进化关系, 并为基因修饰和基因编辑等相关研究打下良好基础。本研究从水稻、小麦、玉米和高粱叶绿体的两组全基因组序列比对结果出发, 查找替换位点, 并统计其侧翼序列的碱基组成, 利用相对熵计算作图展示其特征规律, 并通过相邻特征位点的相对熵贡献分析, 探明核苷酸替换的相邻位点效应。

1 材料与方法

1.1 序列下载与比对

本研究选择4种禾本科作物叶绿体基因组作为研究对象, 序列均从GenBank DNA数据库(https://www.ncbi.nlm.nih.gov/nuccore)中下载得到, 分别为小麦(*Triticum aestivum* L.; GenBank数据库中RefSeq编号: NC_002762)、水稻(*Oryza sativa* L.; NC_031333)、玉米(*Zea mays* L.; NC_001666)和高粱[*Sorghum bicolor* (L.) Moench; NC_008602]。根据禾本科植物进化树(Zheng等2007), 4种作物中水稻和小麦亲缘关系较近, 高粱和玉米亲缘关系较近, 因而将序列数据分为两组, 用MAFFT (https://mafft.cbrc.jp/alignment/server/) (Nakamura等2018)进行序列比对: 小麦和水稻叶绿体基因组的比对以玉米作为外群, 玉米和高粱叶绿体基因组的比对以水稻作为外群。由比对结果分析替换位点和替换类型。外群的选择标准为: 在禾本科进化树中与两种比对作物的亲缘关系较远, 即与两者的最近共同祖先处于不同分支。

1.2 查找替换位点及相邻位点信息

利用C语言编写程序查找替换位点, 如第一条序列某一位点碱基为A, 而第二条序列与外组序列的对位位点均为C, 则将其视为在第一条序列发生了C→A的核苷酸替换, 以此推广至每条序列都可能发生的12种替换。查找替换位点的同时, 记录替换位点上下游序列信息。将发生替换的位点标记为“0”位, 5′端的相邻第一位标记为“−1”位, 3′端的相邻第一位标记为“+1”位, 以此向前后延伸(Ma等2010)。本研究主要记录了替换位点前后各20位核苷酸的信息, 从而得到替换相邻的每个位点所对应的碱基组成。

1.3 相对熵计算与作图

以碱基组成与背景序列差异相对熵作图。相对熵定义为:

$$H(P\,\|\,Q)=\sum P(x_i)\log_2\frac{P(x_i)}{Q(x_i)} \tag{1}$$

其中, $P(x_i)$和$Q(x_i)$分别表示在不同环境下随机变量x各种状态出现的概率, 相对熵[$H(P\|Q)$]则可以来度量P和Q两种概率分布之间的差异(Ma等2010)。在本研究中, P为所研究位点的核苷酸组成分布, Q为背景序列核苷酸组成分布。背景序列即: 如果所研究的替换类型未发生替换, 其前后相邻20个位点的序列。例如对于C→A的替换, 将所研究基因组的C全部找出, 并记录其前后相邻20个位点的序列信息, 将此作为背景序列, 计算其碱基组成分布。由公式(1), 每个位点的相对熵都是由4个

碱基的单独计算结果求和得到, 而每一个碱基的$P(x_i)\log_2\frac{P(x_i)}{Q(x_i)}$值可以看作其对于整体相对熵的贡献。

1.4 阈值的计算

为了确定替换的相邻位点中有哪些位点的碱基组成显著偏离于基因组的平均碱基组成, 本研究引入阈值的计算。按照替换位点上下游序列的碱基组成产生数条40 bp随机序列, 并计算其核苷酸频率作为公式(1)中的P分布; 按相对熵计算中背景序列的碱基组成产生数条40 bp随机序列, 并计算其核苷酸频率作为公式(1)中的Q分布。从而计算得到相对熵, 并将如上过程重复10 000次, 将相对熵的计算结果由大到小排列, 取第10个相对熵值作为阈值, 将实际计算得到的位点相对熵与阈值比较以判断该位点的碱基组成是否显著偏离于背景序列碱基组成。

2 实验结果

2.1 序列比对与替换位点

两组叶绿体基因组全序列比对部分结果如图1所示, 其中阴影部分即为1个核苷酸替换位点。可以发现在高粱、玉米和水稻的比对结果(图1-A)中, 高粱叶绿体基因组的1个位点为G, 而玉米和水稻叶绿体基因组的对应位点为A, 这样可以判断在该位点, 高粱叶绿体基因组发生了A→G的替换。同样, 对于小麦、水稻和玉米的比对结果(图1-B), 可以判断小麦叶绿体基因组在阴影部分所示位点发生了T→G的替换。

2.2 替换位点及侧翼序列统计

基于序列比对结果, 采用编程统计替换位点数目和侧翼序列碱基组成, 最终在高粱、玉米、小麦和水稻叶绿体基因组中分别发现了416、627、3 262和2 396个替换位点。高粱与玉米的亲缘关系较近, 故替换位点与小麦和水稻相比较少。

将替换位点分为12种替换类型, 分别查找统计其上下游各20个位点的碱基组成, 并将其与背景序列的碱基组成进行比较, 计算相对熵值。

2.3 相对熵作图及分析结果

水稻与小麦叶绿体基因组替换位点相邻序列相对熵分析结果如图2所示。在A→C、A→G、A→T、T→C等替换类型中, 两物种作图有很多相似规律和重合位点。大多数替换的相邻+1或−1位点的相对熵值高于阈值(图中水平线所示, 实线对应小麦, 虚线对应水稻), 且高于距离较远位点的相对熵值, 尤其表现在A→G、A→T、T→A、T→C、T→G等替换类型。但高粱与玉米叶绿体基因组比对结果的相对熵作图较为杂乱(图3), 仅在A→G、G→C和T→C等少部分替换类型的部分区域表现出类似小麦和水稻的特征。

2.4 相邻位点相对熵的贡献差异

通过相对熵计算作图已经发现, 在水稻和小麦叶绿体基因组大多数替换类型中, +1与−1位点的相对熵相对较高。为研究CpG高甲基化效应相

图1 四种禾本科作物叶绿体全基因组的分组序列比对部分结果

Fig.1 Part of comparison results regarding the sequence alignment of chloroplast genomes of four Poaceae species

A: 高粱(NC_008602)、玉米(NC_001666)和水稻(NC_031333)叶绿体基因组的比对结果; B: 小麦(NC_002762)、水稻(NC_031333)和玉米(NC_001666)的叶绿体基因组比对结果。左侧NC开头的编号为序列所属基因组的RefSeq编号, 椭圆阴影部分表示发生核苷酸替换的位点。

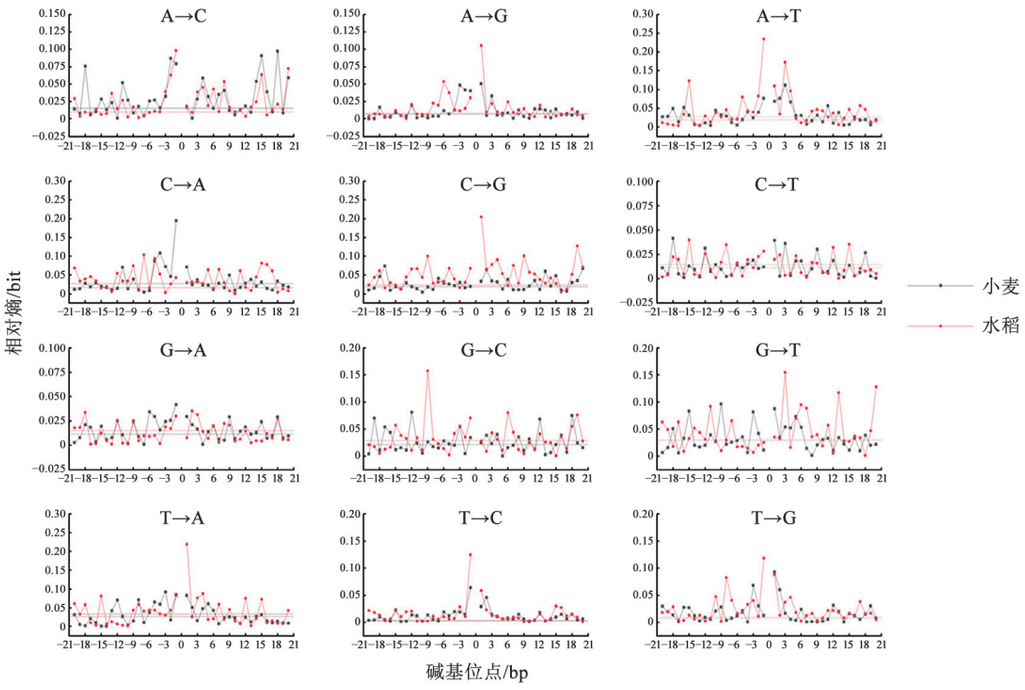

图2 小麦和水稻叶绿体基因组12种替换类型侧翼序列的相对熵

Fig.2 Relative entropy of flanking sequences regarding 12 substitution categories in wheat and rice chloroplast genomes

图中水平虚线分别表示小麦(黑色)和水稻(红色)叶绿体基因组背景序列的相对熵阈值。

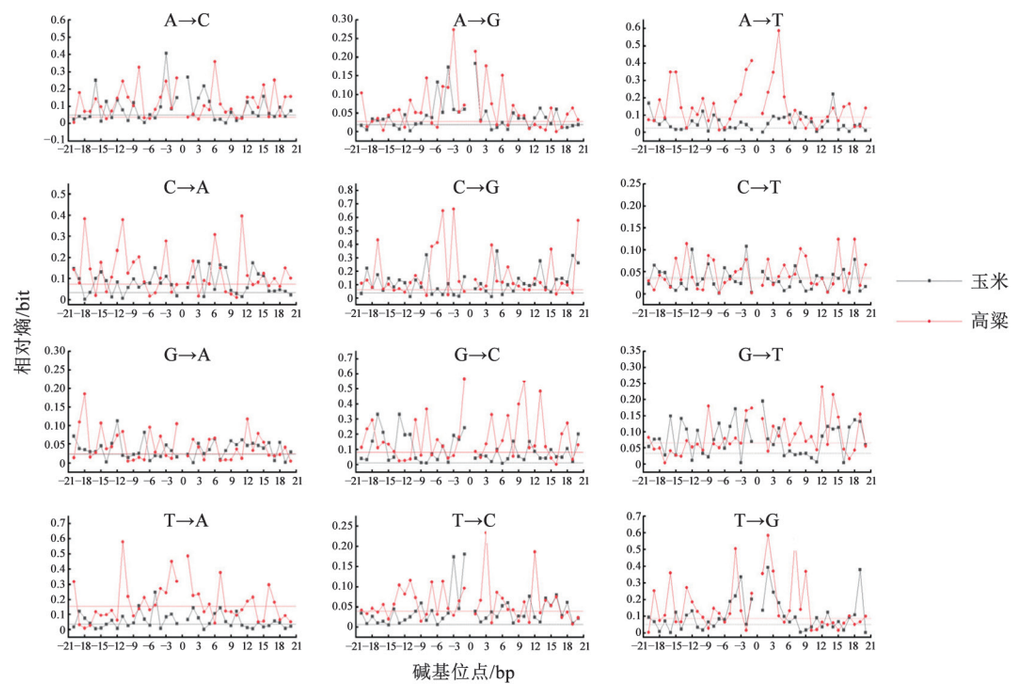

图3 玉米和高粱叶绿体基因组12种替换类型侧翼序列的相对熵

Fig.3 Relative entropy of flanking sequences regarding of 12 substitution categories in maize and sorghum chloroplast genomes

图中水平虚线分别表示玉米(黑色)和高粱(红色)叶绿体基因组背景序列的相对熵阈值。

关位点以及图2和3相对熵值普遍较高的相邻位点, 选取C→T的+1位点、G→A的−1位点、A→G的+1位点、C→A的−1位点、T→A的+1位点和T→C的−1位点对4个物种进行相对熵贡献分析, 结果如图4所示。对于4种禾本科作物叶绿体基因组, C→T的+1位点G的相对熵贡献值最高, 意味着其碱基频率比背景序列高, 即CpG二联子易于发生CpG→TpG的替换, 这一结果符合CpG甲基化效应。相对而言, G→A的−1位点C贡献相对最高, 这是由于作为CpG→TpG替换在互补链的结果(CpG→CpA), 这同样证实了CpG甲基化效应的存在。除此之外, A→G的+1位点G贡献最高, C→A的−1位点T贡献最高, 这说明ApG二联子倾向于发生到GpG的替换, 而TpC二联子倾向于发生到TpA的替换。T→A的+1位点A和G两种嘌呤核苷酸贡献偏高; 而T→C的−1位点C和G贡献偏高, T贡献较低。在4种禾本科作物的叶绿体基因组中替换的相邻位点相对熵贡献表现出了相似的规律, 即使替换数目较少的高粱和玉米叶绿体基因组, 也出现了符合CpG效应和其他二联子的替换规律。

3 讨论

本文采用相对熵的研究方法, 较为直观地分析了4种禾本科作物叶绿体基因组中替换的侧翼序列特征和相邻位点效应。小麦和水稻叶绿体基因组中, 相邻序列的碱基组成受到替换的影响随着距离的增大而减小, 这与哺乳动物单核苷酸多态性(single nucleotide polymorphism, SNP)相邻位点效应(Zhang和Zhao 2004)以及灵长类动物基因中替换相邻位点效应(Ma等2010)的研究结果相类似。对人类基因组的研究发现, 核苷酸替换不是仅仅受到前后两个相邻位点的影响, 而是上下游2~5 bp的序列与相邻位点共同作用(Nevarez等2010)。通过分析人类基因组250万个SNP位点(Zhao和Boerwinkle 2002)以及小鼠基因组等位基因替换位点(Zhang和Zhao 2004), 均发现核苷酸替换的邻近位点受到替换的影响要显著大于更远的位点, 并且随着与替换位点距离的增加而逐渐降低。从图2的A→G、A→T、T→C等替换类型中也可以看出, 离替换越近的位点其碱基组成受到替换的影

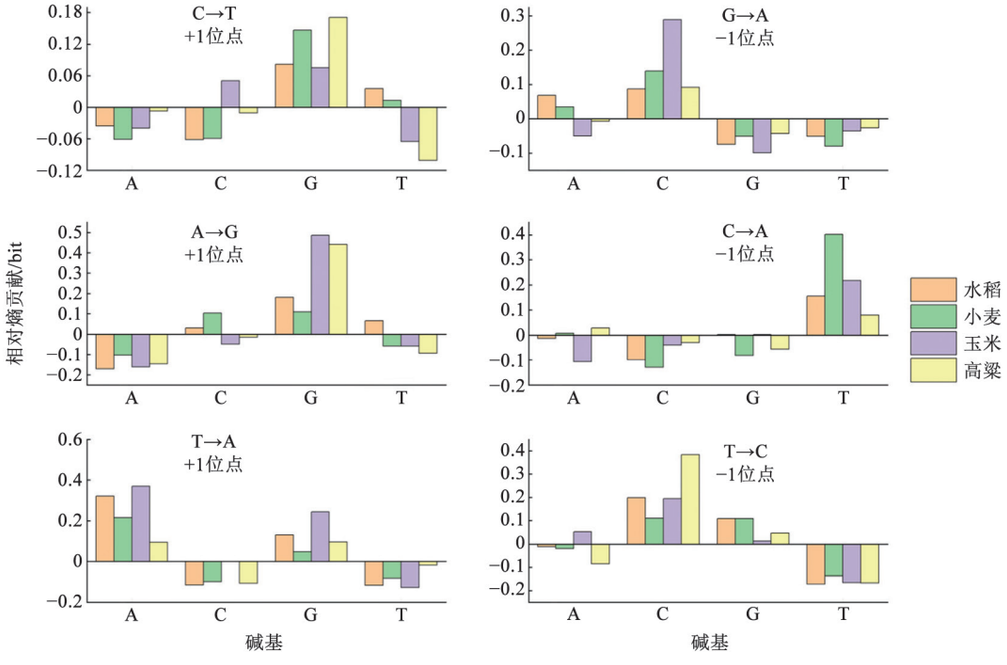

图4 四种叶绿体基因组中部分替换相邻位点的相对熵贡献

Fig.4 Relative entropy contribution of partially substitutive neighboring sites in four chloroplast genomes

响越大, 尤其表现在上下游2~6 bp的序列上。但高粱与玉米叶绿体基因组比对结果的相对熵作图并没有展示出类似于水稻、小麦叶绿体以及动物基因组的规律, 且较为杂乱。究其原因, 高粱和玉米亲缘关系较近, 故产生的核苷酸替换位点数目较少。这种差异导致高粱和玉米叶绿体基因组中替换位点的上下游序列碱基组成随机性更强, 与背景序列差异较大, 因而在某些位点上出现远超过阈值的假阳性。

前人研究发现, 灵长类动物基因中替换的邻近序列有三碱基周期趋势(Ma等2010), 而本研究对于叶绿体基因组的分析并没有发现类似三碱基周期的趋势。这种周期性的产生原因可能是同义密码子的使用偏好。较早的研究对细菌全基因组的相对熵作图也发现了某些物种的部分替换类型具有类似的三碱基周期趋势, 但在大部分替换的侧翼序列并不明显(黄卓然2014)。对于灵长类动物基因组的研究对象是编码区序列, 而本研究的研究对象为叶绿体, 因起源于光合细菌, 故其基因组与细菌基因组结构类似。虽然细菌和叶绿体基因组的非编码序列较少, 但仍然对相对熵计算结果有所影响, 故三碱基周期趋势在很多物种及替换类型中并不明显。

对于相对熵较高的替换相邻位点, 进行不同核苷酸相对熵贡献的研究, 在4种叶绿体基因组中均发现了CpG二联子易于替换为TpG或CpA的CpG甲基化效应, 以及颠换相邻位点的特征碱基组成规律, 这有助于进一步厘清相邻位点效应的产生原因。对于禾本科作物叶绿体两种编码基因的研究已经发现, 颠换相邻位点的A和T含量较多, 且−1位点多为嘧啶核苷酸, +1位点多为嘌呤核苷酸(Morton和Clegg 1995)。本研究中选取的特征替换C→A和T→A均为颠换, 其中C→A的−1位点T较多, 而A、G和C均低于背景序列组成, T→A的+1位点嘌呤核苷酸较多, 嘧啶核苷酸较少。此外, 对于C→T、G→A、A→G、T→C四种转换, 均发现了与颠换相反的结论, 即相邻位点C和G含量较高。这些结果印证了Morton和Clegg (1995)的研究结论, 且可将之扩展到叶绿体全基因组中。

本研究所用的材料尚有局限性, 无法反映禾本科植物整体的进化规律, 随着更多禾本科植物叶绿体基因组的测序完成, 在将来的研究工作中, 可以将相对熵的方法应用到多物种核苷酸或氨基酸序列的分析当中, 为生物信息学研究提供更有效的工具。

参考文献(References)

Chen Y, Zhang Z, Zheng J, et al (2017). Gene selection for tumor classification using neighborhood rough sets and entropy measures. J Biomed Inform, 67: 59–68

Coulondre C, Miller JH, Farabaugh PJ, et al (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. Nature, 274 (5673): 775–780

Dong WL, Wang RN, Zhang NY, et al (2018). Molecular evolution of chloroplast genomes of orchid species: insights into phylogenetic relationship and adaptive evolution. Int J Mol Sci, 19 (3): 716

Huang Z (2014). Patterns of nucleotides flanking sites in bacterial genome (dissertation). Yangling: Northwest A&F University, 11–19 (in Chinese with English abstract) [黄卓然(2014). 细菌基因组中核苷酸替换侧翼序列特征研究(学位论文). 杨凌: 西北农林科技大学, 11–19]

Kessler M, Hoffmann K, Fritsche K, et al (2019). Chronic *Chlamydia* infection in human organoids increases stemness and promotes age-dependent CpG methylation. Nat Commun, 10: 1194

Krawczak M , Ball EV , Cooper DN (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet, 63 (2): 474–488

Liu B, Du Q, Chen L, et al (2016). CpG methylation patterns of human mitochondrial DNA. Sci Rep, 6: 23421

Liu J, Xu F (2003). Constructing phylogenic trees based on relative entropy theory. Acta Sci Nat Univ Pekinensis, 39 (Z1): 76–81 (in Chinese with English abstract) [刘军, 许甫荣(2003). 基于相对熵原理构建生物进化系统树. 北京大学学报(自然科学版), 39 (Z1): 76–81]

Long H, Sung W, Kucukyildirim S, et al (2018). Evolutionary determinants of genome-wide nucleotide composition. Nat Ecol Evol, 2: 237–240

Ma L, Zhang T, Huang Z, et al (2010). Patterns of nucleotides that flank substitutions in human orthologous genes. BMC Genomics, 11: 416

Mader M, Pakull B, Blanc-Jolivet C, et al (2018). Complete chloroplast genome sequences of four Meliaceae species and comparative analyses. Int J Mol Sci, 19 (3): 701

Morton BR, Clegg MT (1995). Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. J Mol Evol, 41 (5): 597–603

Nakamura T, Yamada KD, Tomii K, et al (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics, 34 (14): 2490–2492

Nevarez PA, DeBoever CM, Freeland BJ, et al (2010). Context dependent substitution biases vary within the human genome. BMC Bioinformatics, 11: 462

Panchin AY, Makeev VJ, Medvedeva YA (2016). Preservation of methylated CpG dinucleotides in human CpG islands. Biol Direct, 11: 11

Poelwijk FJ (2019). Context-dependent mutation effects in proteins. In: Sikosek T (ed). Computational Methods in Protein Evolution. New York: Humana Press, 123–134

Shinozaki K, Ohme M, Tanaka M, et al (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J, 5 (9): 2043–2049

Sprouffske K, Aguilar-Rodríguez J, Sniegowski P, et al (2018). High mutation rates limit evolutionary adaptation in *Escherichia coli*. PLoS Genet, 14 (4): e1007324

Sun J, Dong X, Cao Q, et al (2019). A systematic comparison of eight new plastome sequences from *Ipomoea* L. PeerJ, 7: e6563

Tsuruta S, Ebina M, Kobayashi M, et al (2017). Complete chloroplast genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: comparative genomics and evolution of the *Saccharum* complex. PLoS ONE, 12: e0169992

Yang Z, Wang G, Ma Q, et al (2019). The complete chloroplast genomes of three Betulaceae species: implications for molecular phylogeny and historical biogeography. PeerJ, 7: e6320

Yuan T, Li JM, Zhou AP, et al (2015). Analysis of phylogenetic relationship of *Populus* based on sequence data of chloroplast regions. Plant Physiol J, 51 (8): 1339–1346 (in Chinese with English abstract) [员涛, 李佳蔓, 周安佩等 (2015). 基于叶绿体片段序列的杨属系统发育关系分析. 植物生理学报, 51 (8): 1339–1346]

Zhang F, Zhao Z (2004). The influence of neighboring-nucleotide composition on single nucleotide polymorphisms (SNPs) in the mouse genome and its comparison with human SNPs. Genomics, 84 (5): 785–795

Zhang XY, Xin LP (2017). Research progress in chloroplast iron transport proteins. Plant Physiol J, 53 (1): 9–16 (in Chinese with English abstract) [张秀月, 印莉萍(2017). 叶绿体铁转运蛋白的研究进展. 植物生理学报, 53 (1): 9–16]

Zhao Z, Boerwinkle E (2002). Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. Genome Res, 12 (11): 1679–1686

Zheng T, Ichiba T, Morton BR (2007). Assessing substitution variation across sites in grass chloroplast DNA. J Mol Evol, 64 (6): 605–613

Patterns of nucleotide substitutional flanking sequences in chloroplast genomes of four Poaceae species

HUANG Zhuo-Ran, WU Xiao-Min, ZHANG Hui-Jun, DUAN Yong-Bo, ZHANG Qiang*

School of Life Sciences, Huaibei Normal University, Huaibei, Anhui 235000, China

Abstract: The study on the neighboring sites and flanking sequences of nucleotide substitutions not only contributes to exploring the phylogeny of different species, but also lays a basis of the gene editing and modification technique. As suggested by previous studies, CpG methylation effect exists in the mammalian and bacterial genomes, while the base composition of neighboring sites is highly correlated with transversional neighboring sites in the chloroplast genome. Herein, according to the alignment of chloroplast genomes, wheat (*Triticum aestivum*), rice (*Oryza sativa*), maize (*Zea mays*) and sorghum (*Sorghum bicolor*) were divided into two groups. On this basis, substitution sites were found and the base composition of substitutional flanking sites were counted to calculate the relative entropy. Relative entropy mapping shows that the impact on the base composition of flanking sequences decreases with the distance to the substitution sites in wheat and rice chloroplast genomes, while this rule is not found in chloroplast genomes of maize and sorghum. Based on the nucleotide contribution of neighboring sites with higher relative entropy, CpG methylation effect and the special nucleotide composition of transversional neighboring sites were observed in the chloroplast genomes of four Poaceae species. This paper provides a reference for the application of relative entropy into the molecular evolution study of the plant genome.

Key words: Poaceae crop; chloroplast genome; substitution site; flanking sequence; relative entropy

Received 2019-03-26 Accepted 2019-06-17

This work was supported by the Natural Science Funds of Education Department of Anhui Province, China (KJ2016B011), the Program of Domestic Study for Young Scholar by Information College of Huaibei Normal University (gxgnfx2018084), and the Anhui Provincial Natural Science Foundation (1608085MC49).

*Corresponding author (zhangq@foxmail.com).