

# 地钱基因组揭示陆生植物进化过程

翻译自：

Bowman et al., 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171:287-304

译者：

浙江大学农业与生物技术学院应生专业本科生（2015 级）

第一部分：【INTRODUCTION】【RESULTS】【Structural Genomics and Annotation】

成员：丁菁雯、杨桂花、俞文婕

第二部分：【RESULTS】【Insights into Land Plant Evolution】【Transcriptional Regulation】

成员：雷晨旭、朱张士昌、刘博文、肖云鹏

第三部分：【RESULTS】【New Features Adaptive to Life on Land】

成员：龚婉欣、谢子璠、李冰清

第四部分：【DISCUSSION】

成员：袁雯馨、周业凯、郑姚颖、汪玉清

\*该译文来自《生物信息学》课程作业（樊龙江主讲）

## 【摘要】

陆地植物群的进化改变了陆地环境。陆地植物从共同的祖先轮藻进化而来，遗传了轮藻在发育、生化和细胞生物学方面的特性。此外，增添了对陆地在生化和生理上的适应和多细胞单倍体与二倍体世代之间交替的生活周期（促进了耐旱孢子的有效分散），进而演化成了原始陆地植物。我们分析了地钱 *Marchantia polymorpha*（一种基部陆地植物系）的基因组。相对于轮藻，陆地植物基因组编码了一些特殊基因，构成了一些特异生化通道和新的植物激素信号通道（如生长素），拓宽了信号途径，增加了一些转录因子家族多样性。与其他陆地植物相比，*M. polymorpha* 在大多数调节通道上显示出了低遗传冗余性，这与其他古老陆地植物类似。

## 【前言】

陆地植物从原始轮藻进化了 450 百万年。现存的轮藻特征暗示至少有一些陆地生理适应是先于陆地植物进化的，一些陆地植物的基因家族源于轮藻祖先。相比之下，多细胞二倍体孢子体世代（陆地植物的一个特征）可以使耐旱孢子充分分散，对陆地植物逐渐对陆地的支配具有重大帮助。

三种最早分化现存陆地植物系（苔纲、藓纲、角苔纲）缺乏维管组织和真正的根，但是拥有所有陆地植物进化所需的关键创新：多细胞二倍体孢子体、配子体的顶端分生组织（顶端细胞产生三维组织）、孢子体的顶端分生组织和细胞组织分化（提供了形态和生理上对陆地的适应）。苔藓植物系统发育关系仍然是个谜。但是，苔类植物的化石

先于藓类和角苔类，这表明最早的陆地开拓者携带有苔纲的特性。因此，苔纲可能存留更多祖先特性。苔纲是单系发育的，带叶状体的配子体占优势，孢子体退化；具有膜结合的油体和腹侧单细胞假根；没有气孔。苔纲显示了较低的染色体进化，地钱纲（*Marchantiopsida*）的分子进化与其他苔纲植物相比更低。雌雄异体的苔纲有性染色体，可能是群体的祖先。

*Marchantia polymorpha* 是一种地钱，有复杂的叶状体配子体。其化石可追溯到二三叠纪时期，推测该复杂叶状体可以适应干旱环境。由于在实验室生长和基因操作的方便以及存在大量的历史文献中，我们选择了 *M. polymorpha* 作为一种代表性苔类，对其基因组进行了分析。

## 【结果】

### 【结构基因组学和注释】

通过全基因组鸟枪测序法测出 *M. polymorpha* subspecies *ruderalis*（地钱亚种）的核基因组和细胞器基因组。我们测了一个单克隆雌性基因组：由 X 染色体被引入到 Tak-1 常染色体背景的雌性去回交 4 一个已知 Y 染色体序列的雄性 Tak-1 线衍生出来。核基因组大约由 2957 个支架组成，大小为 225.8Mb。*M. polymorpha* 的质体和线粒体序列与之前已经发表过的不同，之前发表的序列来自于粗裂地钱，而不是地钱。细胞器基因组结构和以前的一样，但是进化和功能的结论不一致。

#### 1、基因数量

注释显示了 19138 个编码蛋白基因和 5385 种可供选择的蛋白质编码转录本。为了方便注释，我们获取了代表无菌生长的植物各组织的转录组，以及两种“野生”样本，其某些基因的表达被检测发现无菌生长植物基因的表达中不存在。孢子体和精子器托组织的转录组比萌芽孢子和颈卵器托组织更具有专一性。

在植物的光谱中观察到的，由于基因的翻译选择，高表达的基因密码子的 3 号位更偏好同义嘧啶。*M. polymorpha* 中平均非编码区和内含子的长度均大于其他植物。与研究的其他陆地植物相比，50 个非编码区含有较高频率的 ATG，具有潜在编码 uORF（上游开放读码框）能力，这暗示了翻译组件是如何真正地辨认 *M. polymorpha* 的起始密码子。

核基因组编码 769 个 tRNA 基因（51 个假基因）和 301 snRNA 基因。在 265 个地钱 MIR 基因中，264 个 miRNA 前体（pre-miRNA）在不同的基因组中（包括编码蛋白质的基因）被发现证实。42 个 miRNA 前体存在 ORF。miRNA 前体的表达形式与它们的靶标互动很弱，并且发现了 DCL1 同源序列（miRNA 的加工所需）在地钱和一些轮藻中。

#### 2、基因与基因组进化

*M. polymorpha* 的 6404 个基因在真核同源组（KOG）中找到，12842 个被分配到 OrthoMCL（6348 个基因重叠），另外 419 个基因分配到 PANTHER/Pfam category，留下 5821 个基因没有任何注释。通过 Viridiplantae 的比较分析发现了原始的链形植物与原始陆地植物的同源组的大幅增加。增加类群取样、增加藻类和陆生植物共有的同源序列的分析或许可以解决苔藓植物系统发育问题。KOG 只在 *M. polymorpha* 中有发现，在其他的陆地植物中没有，并且 KOG 往往与真菌基因同源（一致）或与移动元素有关，暗示基因水平转移。通过大量的孤独基因，家族特异或进化相近的基因的基因组测序，可能可以反映地钱与其他物种的系统发育距离（遗传距离）。

与其他陆地植物相关的、在 *M. polymorpha* 中过多表现的多个基因家族可以编码转运蛋白（如磷酸铵）。在 *M. polymorpha* 中丢失的基因家族包括了那些为丛枝菌根的成功定植化需要的，尽管它们会在轮藻和密切相关的地钱属中出现。相关的地钱属形成菌根关联，但是在 *M. polymorpha* subsp. *ruderalis* 中不见了。提高了运输能力，而不依赖菌根体系，这可能是 *M. polymorpha* 在贫瘠干扰的栖息地艰难进行开拓的基因组上的适应。

所有现存的地钱世系的祖先的染色体数目为 9，这意味着古老的全基因组复制的缺失，而且分析证实了 *M. polymorpha* 缺乏原始多倍形。大多数调节基因也为这一假说提供了进一步的支持，有助于证实 *M. polymorpha* 的同源基因水平是处于其他的陆地植物的下端。然而，局部串联排列的基因 TAG 并不少见，以 2 - 11 相邻的旁系同源基因的排列，占 1125 个基因。TAG 的比例（5.9%）处于在开花植物观测值范围的低端（4.6% - 26%），但高于 *Physcomitrella patens*（小立碗藓 1%）。在 *M. polymorpha* 中，75%的 TAG 被编码在同一条链上，而与 *Physcomitrella patens* 相比，小立碗藓由于同源重组的损失，TAG 大量减少。

### 3、重复 DNA

重复序列占了 *M. polymorpha* 常染色体基因组的 22%，这个值比 *P. patens* 小立碗藓低(48%)，但高于一种金鱼藻（角苔纲）（6.98%）。与被子植物相似，长末端重复（LTR）反转录转座子，包括 264 个全长的长末端重复（LTR）反转录转座子，占据了重复序列的最大部分（9.7%）。X 和 Y 染色体的特异重复序列之前有过报道，但没有新增加的性别相关的序列被确认。

### 4、性染色体

*M. polymorpha* 具有性染色体，有一个 X 染色体的‘feminizer 女性化’基因座（基因在染色体上所占的位置），和多个 Y 染色体基因座（为精子运动所需）。先前的 6 MB 的 Y 染色体注释显示有 64 个基因。我们确定了占 4.37 MB 的 9 个 X 染色体骨架，并注释了 74 个 X 染色体上的基因，105 个 Y 染色体基因。X 染色体基因密度（1 基因 57.5 kb）与 Y 类似（1 基因 56.7 kb），大约比常染色体（基因 1 / 11.3 KB）低 5 倍（1/5）。在 74 个 X 染色体基因中，20 个具有在 Y 染色体上有最接近的同源物，它们可以被视为等位基因。这些基因的表达是有关植物生长的，在陆地植物中，木霉或植物界，是保守的，因此，这些染色体代表了原始常染色体（性染色体从中进化）的祖传物。少有证据说明 X 和 Y 骨架之间存在同线性，这表明这些区缺乏重组。进一步，如前所述，X 和 Y 之间的等位基因的同义替换（同义突变）是突变饱和的，反映原始性染色体的起源。X 和 Y 等位基因之间的分歧与在现存的苔类植物（liverwort）的同源序列的分歧是可相互比较的，系统发育分析表明性染色体的起源在苔纲分化之前，与想法“原始的苔类植物（liverwort）可能具有性染色体”一致。

X或Y染色体特异的基因表达时，往往是在生殖器官中特异性表达，符合性染色体进化预测。然而，许多X染色体特异基因位点的基因或基因片段与常染色体的旁系同源基因密切相关，表明它们最近地迁入X染色体，一个类似的模式已经在褐藻报道。因此，只有少数X染色体特异基因存在功能性，在这些少数X染色体特异基因里，一半只有在孢子体中才可检测性地表达，几乎没有明显的‘feminizer女性化’基因座的候选基因。Y染色体特异性基因表现出类似的模式，但包括了几个明显的“motility能动性”候选基因。

## 【陆地植物进化机制】

在剩下的部分中，我们通过将地钱多态基因组含量与轮藻和其他陆地植物的基因组和转录组进行比较，揭示了陆地植物两个方面的演变。首先，我们仔细鉴别了可以区分陆地植物和藻类亲属的基因组特征，并且尽可能描述了他们的起源。起源的推断以类似于化石记录的方式进行，就是说让它们能代表最新的可能日期，除非直系同源藻类序列已经可以精确地定义起源时间。其次，我们注意到在可能的情况下我们可以从系统发育分析推断出预测的祖传陆地植物基因组成分。

## 【转录调控】

地钱基因组包含 394 个（387 个位于常染色体，4 个具有 X 和 Y 等位基因，3 个 X 特异性基因）基因。这些基因分别位于 47 个转录因子（TF）家族中。这些 TF 家族存在于其他陆地植物中，而且除了 Viridiplantae 家族之外，没有其他家族存在于其他真核生物中。地钱的 TF 家族基因占蛋白质编码基因的 2.1%，低于其他陆地植物，高于藻类。这证实了 TF 家族基因数值随着生物复杂性的增加而增加。通过对 18 个 TF 家族的系统发育分析，我们估计出陆地植物祖先中 TF 含量。地钱中的 TF 家族基因含量与陆地植物祖先预测的 TF 家族基因含量相似，具有特殊的谱系特异性基因扩增。

在链形植物早于克里藻的分化和多细胞性演变的基础上，TF 家族基因的多样性开始增加。演化过程中只有一个 TF 家族与陆地植物（GeBP）伴随演变，而且只有很少的几个家族在陆地植物（YABBY, VOZ, ULT）内发展。因此，新的 TF 家族的起源本身对于陆地植物的演化来说并不重要。然而，许多 TF 家族（如 bHLH, NAC, GRAS, AP2/ERF, ASL/LBD 和 WRKY）基因多样性增加表明它们可能对于植物陆地化有帮助。例如，有 14 个 bHLH 亚家族基因仅在陆地植物中发现。进一步研究发现，TF 家族中一些亚家族基因可以直接参与同陆地植物细胞类型相似的细胞类型的形成（被子植物根毛和苔藓植物根茎），该细胞类型具有营养和水分吸收功能。其他 bHLH 亚家族基因在维管植物组织或细胞（例如脉管系统，气孔）的分化中起作用，而在地钱中没有发现。这表明在陆地植物多样化期间预先存在的调节模块中存在联合选择。

对 18 个 TF 家族的系统发育分析揭示了四个不同的反复进化模式，本文以 MYB TF 家族的支派为例（图 4）即：（1）预测陆地植物的分枝；（2）在陆地植物祖先中进化的进化枝，其中地钱基因与许多其他陆地植物旁系同源物是直系同源物；（3）不含被子植物直系同源物的地钱基因进化枝；以及（4）在地钱中表现出谱系特异性扩增的进化枝。模式 2 的示例是进化枝 1,4 和 5（图 4），其成员在调节次级代谢中起作用。图案 4 以 3R-MYBs 为例。

模式 3 在 RWP-RK 和 TALE 类同源结构域家族中是显着的，与被子植物相比，它们在地钱中（和其他基础土地植物谱系）的系统发育上更多样化。与之相反的一个模式的例子是 MADS 盒基因家族，它的 II 型 MADS 盒基因在陆生植物中多样化，但 I 型基因在地钱中是缺失的，尽管它们在其他土地植物谱系中有被发现。

### 1、转录因子在孢子体中的表达

陆生植物的一个定义特征是它有多细胞孢子体。从祖先植物的单细胞受精卵到多细胞孢子体的进化可能是通过现有和新的基因调控网络的建立造成的。在两个相关的苔藓中，大部分重叠的 TF 主要是孢子体表达。在地钱中，41 个 TF 主要在孢子体发育过程中表达，但只有 10 个基因是类似苔藓基因的潜在直系同源物（表 S7）。其中有两种是已知可调节植物繁衍的 TALE-HD 基因。地钱与苔藓孢子体基因最小的重叠可能反映了在地钱和苔藓之间的孢子体解剖学和形态学的显著差异。

### 2、染色质

地钱的染色质和其他的真核生物中大多数的染色质相关基因属于同系物（表格 s7），

在染色质结构和修饰中起作用的一部分基因在雄性生殖（33%）或孢子体组织（19%）中高度表达。地钱染色质基因中，在具有多个旁系同源物的家族（18个中的13个家族）中经常观察到独特的孢子体表达，这表明染色质在配子体到孢子体转换期间发生了变化，例如 *P.patens*。这种染色质相关的基因也可以解释特定的孢子体特征，例如二倍体细胞中的有丝分裂或者可能需要的细胞印记，又或是剂量补偿。类似地，染色质皱缩的规律可能解释雄性配子形成过程中存在的特定表达模式。

DNA 甲基化是真核生物表观遗传的修饰方式，它在一些蛋白质的编码基因的部分片段的沉默也调控中发挥着作用。地钱营养配子体显示地钱中的 DNA 甲基化非常多而且在 23-24nt sirna 簇在基因组基因座有重复片段，这匹配维管植物中的模式，但不超过基因体，除了性染色体上的基因，它的甲基化可能从附近的重复片段转播而来。除了不存在 *cmt3* 基因之外，地钱具有与其他陆地植物相似的一套甲基转移酶，这可能与地钱缺乏基因体甲基化有关。

### 3、RNA 生物学

我们在地钱的单拷贝极影中发现编码介导 mRNA 剪接功能蛋白的 SR 家族蛋白的基因。PPR 蛋白指导在 RNA 细胞器中的加工步骤，植物特异性 PLS 类 ppr 蛋白具有介导 RNA 编辑。地钱中单个 PLS 基因缺少其他 RNA 编辑蛋白的功能域，这与地钱细胞器中的 RNA 编辑功能丧失一致，表明地钱中 PLS 蛋白与其他陆地植物的功能不同。

### 【信号通路分化】

多细胞生物的标志是种类众多的信号通路，细胞以此交流，并且它影响细胞特化、分化和生理。在大多数情况下，地钱的基因组编码一套微小但是完整的陆生植物信号分子网络。

#### 1、F-box 基因介导的荷尔蒙途径

生长素调节植物生长、分化、对重力和光-热的感应。陆生植物经由吲哚-3-丙酮酸合生长素的保守信号通路，是由一个常染色体的 *MpTAA* 基因和五个 *YUCCA* 并系同源基因（其中有一个在地钱配子体中表达）组成的。在绿藻和轮藻里面都有一个和陆生植物进化枝（clade）[所有的陆生植物]直系同源的基因，包括 *PAA* 和它的并系同源物 *TAR* 基因。但是在轮藻类里面没有发现和陆生植物 *YUCCA* 直系同源的基因，表明这种生长素合成途径是在陆生植物里面进化出来的。

地钱的生长素转录应答装置是由一个 *AUX/IAA* 的直系同源基因、*TOPELESS*（*TPL*）、三类 *ARF* 基因、还有 *TIR1* 受体组成的。在轮藻中至少有两类 *ARF* 家族，都具有[with the association of...]在植物界进化中很早出现的 *B3* DNA 结合和 *PB1* 蛋白质互作结构域。通过 *B3* 结构域的丧失和 *I/II* 结构域（和 *TPL* 辅阻遏物和 *TIR1* 受体加上生长素互作）的出现，*AUX/IAA* 基因从一种远古的 *ARF* 进化而来，带有 *PB1* 结构域的轮藻蛋白和陆生植物缺少结构域 *I/II* 的 *AUX/IAA* 蛋白最为相似。因此，尽管轮藻具有陆生植物 *AUX/IAA* 基因的直系同源基因，它们对应的蛋白却缺少经典的[canonical]生长素转运应答的结构域。在地钱里面，多了两个基因（*MpAXI1*, *MpAXI2*）编码 *PB1* 结构域但是缺少 *PB3* 结构域，这可能是为了提供额外的 *ARF* 调节作用。轮藻具有一个 *F-BOX* 基因，它和 *TIR1*（生长素受体）还有 *COI*（茉莉酮酸酯受体）的基因是直系同源基因关系。但是，这些藻类的 *F-box* 直系同源物不具有保守的和相应的陆生植物配体（*JA-Ile/JAZ*、生长素、*AUX/IAA*）相互作用的关键残余，说明它们既不是生长素也不是 *JA* 受体。虽然根据较早之前的假设，生长素和 *JA* 感知机制似乎是通过基因的重复和新功能化在古代陆生植物里面从古代藻类的 *F-box* 进化而来的。

生长素极性运输（*PAT*）在地钱配子体中发生。其依赖于 *PIN* 和 *ABCB* 外流以及 *AUX*

内流促进者[facilitator]。PIN 和 ABCB 直系同源基因都在克里藻属[属于绿藻的丝藻科]里面存在,提高了 PAT 在轮藻中存在的可能性。地钱具有 5 个 PIN 基因与几个 ACAB 生长素外流易化物直系同源基因。一个 AUX 直系同源基因存在于地钱中,但是轮藻中没有识别到类似序列。AGC VIII 组激酶和调节 ACAB 和 PIN 定位和活性的类似 NPH3/RPT2 的蛋白在植物登陆之前就已经出现,但是在古代陆生植物中发生了分化。因此,陆生植物生长素外流易化物和转运蛋白由古代和更晚演化出来的组件组成,暗示着生长素极性运输比陆生植物更早出现并且在陆生植物进化过程中复杂性增加。

植物通过茉莉酸[也叫茉莉酮酸酯]信号通路对生物和非生物胁迫做出反应。与其他苔藓植物类似,地钱可以合成 JA-前体 OPDA (12-氧-植物二烯酸),但缺乏产生维管植物激素 JA-Ile 的 OPR3 (OPDA 还原酶)。尽管如此,COI1 共受体的单直系同源物、JAZ 阻遏物、MYC 转录因子和 NINJA 适配器以及 TPL 联合阻遏物都在地钱中有发现。没有明确的 JAZ 或 NINJA 直系同源基因存在于绿藻或轮藻中。因此,JA 感知和抑制机制 (COI1-JAZ, NINJA) 出现在陆生植物祖先中,它通过新获得的 JA 或 OPDA 相关配体招募了之前存在的 MYC TF 来调节生命活动。

独脚金内酯 (SL) 是抑制芽的分枝和促进丛枝菌根共生的分泌信号分子。苔藓植物 SL 生物合成途径不同于维管植物。地钱中存在一组最小的 SL 信号组分,其中只有一些成分存在于轮藻中,暗示了陆生植物 SL 途径是由先前存在的和新进化出来的组件组成的。同样,一部分维管植物 GA 合成酶的直系同源物在地钱中被鉴定出来,它们和与不存在于苔藓植物中的典型 GA 化合物一致。地钱中存在除 GID1 外的 GA 信号通路组分。因此,与小立碗藓相比,地钱 GA 信号模块更接近维管植物,但是其中缺少 GID1 表明陆生植物的进化是逐步进行的。

## 2、脱落酸 (ABA)

脱落酸 (ABA) 与休眠有关,并且可以让陆生植物能更好的适应潮湿的环境。地钱自身可以产生脱落酸,但生长阶段的胞芽在高脱落酸处理下无法适应干燥的环境。除了有关 x 染色体遗传,代谢,脱落酸的转移这些方面的基因外,脱落酸生物合成酶的单同源基因 (single orthologs) 都能在地钱中找到。PTR 型转移子也不能找到。细胞对于脱落酸的反应是受胞内受体 (PYL) 调控的。在这过程中,ABA—PYL 复合体通过限制 A 组蛋白的磷酸酶 2C (PP2C) 来对 SNF1 蛋白激酶 2 (SnRK2) 起作用。MpPLY1 是一个功能受体,它可以补充拟南芥的 *pyr1pyl1pyl2pyl4* 突变体。地钱有很多不同的 PYL 同源基因 (paralogs), 在其他植物中并没有发现这些同源基因,而在孢子体表达模式中的 MpPYL 同源基因亚功能化也证实了这点。PP2C 和 SnRK2 在绿色植物界中处处可见,但是 PYL 受体的演化时期更接近于现在。因此,对于远古陆生植物来说,PYL 受体的收集是生成脱落酸的关键。下游的 TF ABI3 在苔藓植物耐旱性方面具有关键作用,当 AREB 在植物界中出现时,TF ABI3 就已经和陆生植物一起演化出来了。

## 3、双组份系统荷尔蒙路径 (Two Component System Hormone Pathways)

与 F-box 激素调控路径不同,细胞分裂素和乙烯双组份系统信号转导路径在轮藻中可以找到,但是绿藻中没有,这意味着双组份系统信号转导路径起源于轮藻进化早期。在其他非种子植物中,基因编码一个乙烯形成酶 (ACO), 这种具有种子植物特点的过程并没有出现在地钱中。

## 4、受体激酶信号途径

细胞通过跨膜受体来感应细胞外的分子物质,同时陆生植物相对于绿藻来说会有一个 RLK / Pelle 激酶的激增过程,这个过程会被记录下来。富含亮氨酸的重复受体激酶家族 (LRR-RLK) 是陆生植物 RLK / Pelle 激酶的最大子类。地钱的 107 个基因都有 LRR 和激酶域,其中 15 个子类中有 14 个在拟南芥中也有发现,而如 BR/1 和 PSKR1 等是明显缺失

的。轮藻只有少部分的子类，这预示了远古陆生植物中 LRR-RLK 基因的多元性。拟南芥中，至少有 5 个肽家族起到的是信号的作用。CLE, IDA, EPFL 和 RALF 这些家族的成员在拟南芥中存在，但是我们不能预测其他家族是否存在。在地钱中还发现了能区分肽的受体，但是在轮藻中并没有发现，这意味着这些肽信号通路在远古陆生植物中就有存在。类似的，那些遍及陆生植物的大多数的 RLK/Pelle 激酶在地钱中有发现，而在轮藻中没有发现，这暗示了陆地生物的出现进化在防御信号系统方面都具有多元性。

与 RLK/Pelle 激酶不同，大多数 MAPK(K)(K) 亚科的基因在轮藻或者绿藻中可以找到，这预示了大部分现存的 MAPK 的集联多样性比陆生植物更早出现，这和细胞循环机制一样。同时，所有陆生植物亚科的 PPP 和 PP2C 磷酸酶也已经在轮藻和绿藻中发现。与其他陆生植物相比，地钱中的 MAPK 和磷酸酶机组被最小的同源基因所编码，我们预测陆生植物也是类似的。相反，组氨酸激酶 (HK) 多样性在轮藻进化过程早起就已经出现了，在陆生植物进化之前和期间丢失了一部分。在其他信号通路中，只有 F-box 中的一种起源于陆生植物，而且远古陆生植物的 PEPB 的多样性高于现存的被子植物。

## 【陆生环境适应新机制】

从一个水生或是半水生的状态转变为陆生需要对环境进行适应，包括非生物和生物方面的，需要新的或是改善的生物化学。

### 1、光氧化胁迫适应性

为了适应环境光照条件和进行有效的光合作用，植物进化出了不同的光受体家族，所有的被子植物光受体早于陆地植物的进化。M. polymorpha 具有类似于古老陆地植物所推测拥有的单系同源光受体和核心光信号组分，此外还有一对地钱特有的 LOV 结构域蛋白 (表 S10)。苯丙素类，特别是类黄酮，就像是陆地植物抵挡紫外线的防晒霜，其由 UV-B 通过 UVR8 光受体上调。地钱产生许多小苯丙素代谢物，包括抗菌的联苳、吸收 UV 的黄酮苷、黄甙苷和细胞壁红色素沉积，可能是花青素里的毒素 A 和 B。M. polymorpha 中存在着编码苯丙素通路、莽草酸通路和 UVR8 信号传导的核酸酶，主要是单拷贝数或低拷贝数，除串联排列的 PAL 和 CHS (表 S10)。没有关于藻类中类黄酮的令人信服报道，并且我们也没有发现能够证明轮藻中存在苯丙素类生物合成的基因。这些酶 (例如 PAL) 中的至少有一些可以通过土壤微生物的水平基因转移而得到。

### 2、水份控制

细胞壁硬度和不透性使植物中水传导和保留更为有效。典型的陆地植物细胞壁生物化学特性存在于轮藻类植物中，它们的细胞壁成分和机械性能，反映了其早期对陆地的适应。对地钱中编码细胞壁相关酶的基因进行注释后，表明这种观点是正确的，而且发现大多数的糖苷转移酶和水解酶家族都存在于轮藻中 (表 S10)。地钱中的碳水化合物酯酶，多糖酶，果胶甲基转移酶的多样化，以及木葡聚糖内转糖苷酶/水解酶 (XTH) 的来源与陆地植物的出现相一致。轮藻中 XTHs (木葡聚糖内转糖苷酶/水解酶类) 与木葡聚糖都缺乏。植物中 MpXTH 和扩展蛋白 (该标注表明) 的多样化表明，与种子植物相比，木葡聚糖 - 纤维素网络的重构可能比果胶在调节地钱细胞壁机械性能中更重要。

与水的分布有关的胞间连丝，细胞间的质膜连接在轮藻类或祖先陆地植物中不断地在演化。大多数已知的被子植物胞间连丝的蛋白同源物仅限于维管植物或具有广泛的分布，包括缺乏胞间连丝的物种中也存在；因此，它们的起源早于或晚于胞间连丝的进化；只有两个例外是陆地植物特有的 (表 S10)。在细胞分裂末期，细胞板形成导致胞质分裂时，胞间连丝亦随之发生。这样的变化在 streptophytes (一类绿色的淡水藻类) 的衍

生物中慢慢进化而来而且这种变化和轮藻多样化的膜运输机制有关。

从 *Streptophyte* 进化枝中看，在细胞分裂机制进化的过程中，包括编码与胞质分裂有关的突出融合蛋白 *SYP11/KNOLLE* 在内的 *SYP1* 基因的进化，可能在细胞分裂机制的进化过程中起了重要作用。与其他陆地植物不同的是，地钱也保留了原始的类似于中心体的极点组织者，因此代表了祖先向心分裂和衍生的离心分裂机制之间的过渡状态。

聚合物木质素增加了次生细胞壁的硬度和不稳定性，它的形成是水传导系统和结构支撑的关键创新。

类似于小立碗藓，地钱具有除了阿魏酸-5-羟基化酶(表 S10)外的所有木质素生物合成基因，而我们还没有它们中发现真正的木质素聚合物。因此，根据小立碗藓推断这些酶可能作用于“前木素”合成的途径中。其他陆生植物的水传导细胞的分化是由能增强细胞壁和程序性细胞死亡(PCD)基因表达的 *VNS* 子家族 *NAC TFs* 控制的。虽然地钱没有内部的水传导细胞，但它拥有一套核心的 *vns*-下游基因，这些基因可以在增厚次生细胞壁中发挥作用，也可以下调 *PCD* 的表达。轮藻具有 *VNS*-下游酶的同源性，但缺乏 *NAC* 和 *MYB* 的同源序列 (图 4, 表 S10)，这意味着 *VNS* 的调控网络是建立在原始的陆地植物上的，后来又在衍生系中选择直接指导细胞分化。

地钱叶状体的背侧表面是疏水的，形成了防止水流失的角质层，同时保护植物免受昆虫和紫外线辐射的伤害。陆生植物角质层由细胞外的角质聚合物及散布夹杂在表皮的蜡质组成。地钱中存在调节角质层生物合成和沉积物的基因的同源序列。相比之下，地钱中角质单体生物合成(*CYP86A*)的同源性还没有被确定，地钱纲中的角质单体不同与其他陆地植物。类似地，地钱的蜡质单体的生物合成同源物也与陆地植物不同。

尽管在一些轮藻类植物中可能存在类似的表面层，但在轮藻中并没有发现与之密切相关的角质生物合成或转运蛋白同系物。

聚合物孢粉素是孢子外膜的主要组成部分，目的是保护孢子免受干燥。在孢粉素生物合成过程，通过酰基辅酶 A 合成酶，花药特异性查耳酮合成酶和四甲基  $\alpha$ -吡喃酮还原酶产生羟基化烷基吡喃，地钱具有所有酶的类似物质。植物合成了多种特殊的代谢产物，部分是由于细胞色素 P450 单氧酶(*CYPs*)、2-氧调节依赖的双氧酶(*2OGDs*)和 UDP 依赖的糖基转移酶(*ugt*)产生的。与调节基因不同的是，地钱大多数的 148 个 *CYP*，41 个 *UGT* 和 38 个 *2OGD* 基因代表了一个新的家庭，表明特异性新陈代谢大量的系统特异性多样化。

除了水之外，通道也能促进调控相关的感受以及外源分子的吸收及其在整个植物中的分布。大多数被子植物膜通道和转运体都有与地钱同源的基因 (表 S10)，其中大部分的多样性都是从藻类的祖先中产生的，除了 *NPF* 转运蛋白和 *ABCG* 转运蛋白的一个亚类在祖先的陆地植物时期开始表现出多样性。

### 3、防御系统

信号分子水杨酸 (SA) 调节机体对病原体的防御反应。在 SA 生物合成和信号的所有关键部分的同源基因在地钱中确定 (表 S10)。无论是 PAL-依赖的 SA 生物合成途径还是 SA 的转录反应介质 *NPR1*，只存在于陆地植物。陆地植物的下游 *PR1* (发病相关) 基因家族的扩增可能跟随着 *NPR1* 基因的获得。

植物通过两种受体感受病原体。一种是质膜定位模式识别受体，*FLS2* 和 *EFR*，属于 *LRR-RLK* 亚科 XII 并要求 *SERK* (*LRR-RLK* 亚科 II) 共受体。地钱没有 *FLS2* 或 *EFR* 同源基因 (表 S10)，但其他 *LRR-RLK* 亚科 XII 基因可能在检测细菌的激发子中起作用。与此相反，*SERK* 基因在陆地植物中被保存。另外一种是在细胞内受体，含有核苷酸结合位点 (*NBS*) 和 *LRR* 结构域(*NBSLRR*)，直接或间接地识别病原菌毒力分子。需要 *NBS-LRR* 蛋白活性的分子伴侣复合物的出现早于陆地植物 (表 S10)。*NBS-LRR* 基因与拟南芥 (536) 和小



立碗藓（165）中的大群基因一起快速进化。地钱只拥有 34 个基因，其中一些是成对的，暗示被子植物中有类似的基因组结构。非种子植物中已知的 NBS-LRR 下游信号基因的缺失表明这条途径后来多样化。

苔类植物所特有的油体，有积累萜类化合物和阻止病原体和草食动物的作用。地钱具有 7 种典型植物萜类合酶和 32 种微生物萜类合成酶（MTPSL）基因，包括 10 个部分的基因。在体外实验中，几个地钱 MTPSL 和倍半萜、单萜合酶起一样的功能。祖先 MTPSL 基因是从细菌和真菌的基因水平转移获得的，给苔类提供化学防御来阻止食草动物和病原菌的侵入。

#### 4、基因横向转移

鉴于 MTPSL 的明显水平基因转移（HGT），我们寻找与地钱和真菌的系统发育分布相似的科，除了所有其他真核生物。包括 MTPSL，我们确定了与真菌基因最密切相关的 10 个科（42 个基因）和与细菌基因最密切相关的 5 个科（31 个基因），作为从微生物到地钱 HGT 的候选人（表 S11），与小立碗藓中报告的类似。与子囊菌纲或担子菌纲的序列相似性意味着多个转移不是从菌根真菌复合体获得的，通常在复杂的叶状体苔类中是球囊菌门。尽管（如 Yue 所提出的）从远古的陆地植物中获得的大部分基因在轮藻中出现了，但是考虑到可能的 HGT 频率，更深入的分析可能揭示了远古陆地植物在领土化中接触了新的细菌时的其他事件。

## 【讨论】

### 【基因组构成】

我们的数据与地钱植物谱系中一种二型性染色体的祖先保持一致。

在以二倍体世代为主的生物体中，独特的异型性染色体由于有害突变的积累而遭受退化（Muller）。相反的，以单倍体世代为主的生物体的性染色体的进化路线则完全不同：

（1）性染色体拥有相同的特征，即两条性染色体有相同的退化；（2）退化是被限制的，它保留了配子体基因而丢失孢子体基因；（3）染色体尺寸大小的改变是异染色质的原因。（Bull）

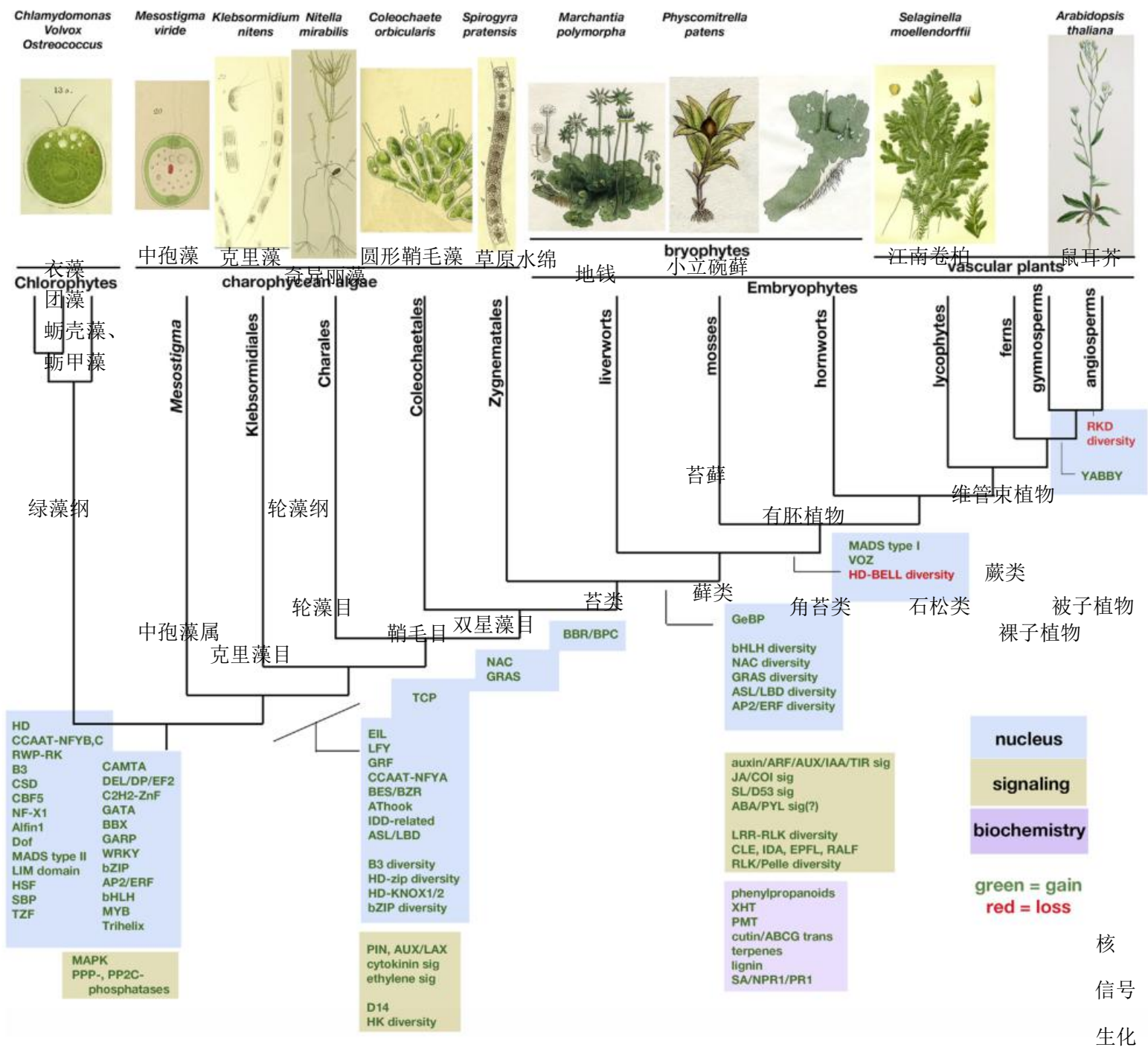
如果地钱的性染色体的祖先常染色体与现存的常染色体相似，则这一理论未支持的方面是（2）有限退化，因为通过性染色体基因密度可以推导出基因倍增时的某些基因缺失。（生物合成等基因留下，调控基因缺失）

相比于其他已测序的陆地植物基因组，地钱基因组的独特之处是其很少存在调控基因的冗余，这暗示着现存植物最近的共有祖先很可能拥有与地钱类似的调控基因组（无论地钱植物的系统发育位置如何）。地钱中缺少重复的调节因子，但它的其他生物合成、代谢以及结构基因却并没有缺失冗余基因。这个情况该怎么解释？从草履虫到被子植物的全基因组复制均导致了转录因子和调控分子冗余的保留。相反地，由于剂量敏感性（dosage sensitivity），小范围转录因子基因的复制被淘汰（因为它们通常在复合体中工作），然而生物合成、新成代谢及结构基因的小范围复制被留下。由于地钱中古老的全基因组复制的证据很少，因此 *M. polymorpha* 冗余基因较少可能是缺乏古老的全基因组复制。（以往研究说明全基因组复制会产生大量冗余基因。）那么问题来了，“为什么地钱中没有古代的 WGD？”因为根据以往的研究（Mable），可以发现地钱绝大多数特征都是有利于多倍体发生的：比如繁殖速度快，自我兼容性好，二分体形成速度快。认为多倍体在动物中比植物中少是因为其具有二型性的性染色体，而这些异型性染色体在减数分裂中的四倍体形成中难以组合成可存活的染色体组合。因为性染色体并不能杜绝多倍体形成（同型性染色体），所以我们猜想早期二型性染色体的进化导致了地钱的基

因稳定性（也就是不形成多倍体）。地钱纲减数分裂中二分体的形成并不罕见（减数第二次分裂着丝点不分离形成的二倍体配子），但祖先多倍体仍然缺失，说明了其对多倍体基因型的筛选排除，可能是由于其性染色体低效的配对（**inefficient pairing of the sex chromosomes**），就像我们在诱导得到的地钱多倍体中均观察到的一样。而现存的地钱，其多倍体植株都不可避免成为雌雄同株。苔藓，蕨类以及被子植物都没有二型性染色体所以都很容易变成多倍体，而角苔类，因为性染色体异型，很少有多倍体存在，初步证实了作者的理论。所以我们猜想，早期地钱家系性染色体的进化导致了 WGD 的缺失，从而减少了调控因子的冗余，使其调控基因和古代陆生植物相像。调控基因的冗余是限制还是促进地钱的进化还需进一步研究，但苔纲与藓纲和角苔类相比的配子体形态学的更大的多样性可能会对研究有一定的帮助。

### 【陆生植物进化】

先前的基因组比较已经开始研究了陆生植物特有的基因组。利用从轮藻到陆生植物的基因组和转录组学资源，我们改进了许多基因家族的进化起源，特别是陆地植物特有的信号通路。与那些的轮藻类亲属相比，陆地植物具有编码新生化途径的基因（角质，苯丙素，萜类，木质素，SA），新的 F-box 介导的植物激素信号通路（生长素，JA，SL），扩增的受体样激酶（RLK/Pelle）及其肽配体（CLE，IDA，EPFL，RALF），但是并不是 MAPK 或磷酸酶家族，而某些 TF 家族更加多样化（图 6）。一般情况下，有新功能化的基因倍增足以驱动分子水平多样化（TF，角蛋白，RLK/Pelle）。然而，在其他情况下，分子多样化还需要新功能化组装进化过程中仍保留和新产生的基因之间的分子互作（*de novo* molecular interactions），比如植物生长素和 JA 的组装（assembly）。由于生长素广泛影响着陆生植物的进化机制，其转录信号通路的演变为陆生植物的形态多样化奠定了基础。最后，陆生生物之间的交互不仅建立了共生关系（例如菌根真菌），同时也获得了能提供新的生化物质（萜类化合物，苯丙素，生长素生物合成）的横向转移基因。这些革新对于陆生植物结构的发展演变和适应新的生物和非生物压力都是至关重要的。



附：陆生植物属性演变（显示了从藻类祖先进行陆生植物演化过程中涉及转录，信号传导和生物化学的基因的获取或损失，其中反映了最新可能的共同祖先的位置）