

技术与方法 Techniques and Methods

一种改良的克隆小麦 *GLP3* 基因启动子的 TAIL-PCR 技术

陈军营, 孙佩, 王德勤, 陈新建*

河南农业大学农学院, 郑州 450002

An Improved TAIL-PCR Method for Cloning of *GLP3* Gene Promoter from Wheat Genomic DNA

CHEN Jun-Ying, SUN Pei, WANG De-Qin, CHEN Xin-Jian*

College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China

摘要: 以小麦基因组 DNA 为模板, 用一种改良热不对称交错 PCR (thermal asymmetric interlaced PCR, TAIL-PCR) 技术克隆到全长为 1 748 bp 的小麦 *GLP3* 基因的上游侧翼序列。测序结果表明, TATA-box 位于基因转录起始位点 CAT-box 上游 -27 bp 处; CAAT-box 位于 CAT-box 上游 -163 bp 处; ATG 位于 CAT-box 下游 94 bp 处, 5'UTR (非编码区) 序列全长 93 bp, 表明该序列为小麦 *GLP3* 启动子序列。

关键词: TAIL-PCR; *GLP3* 基因; 启动子

染色体步行是指以生物基因组或基因组文库中的已知序列为依据逐步探知其侧翼未知序列的技术, 此方法适合于长距离步行, 但必须进行酶切和连接反应, 比较烦琐; 而基于 PCR 方法改进的染色体步行技术比较简单, 但必须在特殊的 PCR 条件下进行, 且无法识别 PCR 产物中“特异条带”(Ochman 等 1988; Rosenthal 和 Jones 1990; Lagerstrom 等 1991; Sarkar 等 1993), 从而影响了这些技术的应用。

Mazars 等(1991)提出热不对称 PCR (thermal asymmetric PCR, TA-PCR) 技术克隆基因侧翼序列; 在此基础上 Liu 和 Whittier (1995) 采用交错改变退火温度的方法改变 PCR 的“严谨性”即热不对称交错 PCR (thermal asymmetric interlaced PCR, TAIL-PCR) 技术, 增加“特异带”的产量, 消除“非特异带”的干扰, 巧妙的解决了“非特异带”的识别问题。其主要原理是: 根据已知 DNA 序列, 分别设计 3 条同向且退火温度较高的特异引物 (specific primer, SP) 和 4 种退火温度较低的兼并引物 (arbitrary degenerate primer, AP), 然后进行 3 次巢式的热不对称 PCR 反应即可获得某基因的侧翼序列。该技术具有简单快速、高效特异、并能对产物直接测序等优点, 已成为分子生物学研究中非常实用的基因侧翼序列克

隆技术 (Benkel 和 Fong 1996; Jones 和 Winistorfer 1992, 1997; Yan 等 2003; Parker 等 1991; Qin 等 2003; Antal 等 2004), 在克隆基因、寻找已知基因的启动子和调控元件、研究基因的内含子和外显子边界、探知特定位点的上游和下游序列、分析转基因生物基因组中外源基因的插入位点等方面已有广泛应用。

但在实际工作中我们感到 Liu 和 Whittier (1995) 的方法仍有需改进的地方, 如第 1 次 TAIL-PCR 前的 10 个降低等严谨度的 PCR 循环可以省去; 3 次 PCR 循环中的变性时间和延伸时间应该延长; 3 次循环中应该都增设不同的退火温度, 这样会更有利于特异目的片段的富集。

小麦 *GLP3* (germin like protein 3) 是小麦萌发过程中的一种蛋白质, 具有 SOD 和 OXO 双重活性, 并且具有加固细胞壁的功能, 与植物抵御生物和非生物胁迫有密切关系 (Caliskan 等 2004; Carter 和 Thornburg 1999; Chen 等 2006; Woo 等 2000; Kim 和 Triplett 2004)。研究表明, Germin

收稿 2007-03-19 修定 2007-05-24

资助 国家转基因植物研究与产业化专项基金 (JY03-B-19-2) 和河南省杰出人才创新基金 (No.022100090)。

* 通讯作者 (E-mail: xinjian@371.net; Tel: 0371-63554960)。

的生理作用是多方面的, 但对其启动子的克隆还未见报道。本文以小麦品种豫麦18的基因组DNA为模板, 采用改良的热不对称交错PCR技术, 成功克隆到了 *GLP3* 基因启动子的全长序列。

材料与amp;方法

1 材料

小麦(*Triticum aestivum* L.)品种‘豫麦18’。

2 提取基因组DNA

选取200粒饱满的麦粒, 置于铺有2层湿润滤纸的发芽盒中, 于室温下发芽。幼苗长至10 cm高时取叶片, 用CTAB法提取其基因组DNA作为扩增未知侧翼序列的模板。

3 引物设计

以已知基因序列为基础设计了退火温度较高(60 °C)的3个特异引物SP1 (GTAGAGCTTGTCC-C-AGAGTCGAG)、SP2 (TCACCTTTCAACAC-GATGCCAATG)和SP3(GCTTGCATGGATG-CCCGTTCACAG), 引物方向为需要扩增的未知区域方向。AP1 [TG(A/T)GNAG(A/T)ANCA(G/C)-AGA]、AP2 [AG(A/T)GNAG(A/T)ANCA(A/T)-AGG]、AP3 [CA(A/T)CGICNGGAIA(G/C)GAA (I: 次黄嘌呤腺苷酸)]和AP4 [TC(G/C)TICGNAC-IT(A/T)GGA]为经过独特设计且退火温度较低(44 °C)的简并引物, T_m 为47~48 °C, 按公式 $69.3+0.41(\%GC)-650/l$ 算出(l 为引物长度)。同时设计测序引物SEQR1 (AGCAAGCTTAGCTAGCACTG)、SEQR2 (TTCAGTTTTAAGAACGATTA)、SEQR3 (TACCCACA ACTATCAATTCA)、SEQR4 (TTAGCATGATCGGCGCATTT)和SEQR5 (CCAGCGAAGCAACGAAAACA)。

4 TAIL-PCR过程(图1)

第1次PCR: 用TaKaRa LA *Taq*TM DNA聚合酶, 以获得的基因组DNA为模板, 分别以SP1/AP1、SP1/AP2、SP1/AP3和SP1/AP4为引物对, PCR扩增目的片段。分别取5 μL PCR产物进行琼脂糖凝胶电泳检测, 其余作为第2次PCR的模板。

第2次PCR: 以第1次PCR产物为模板, 分别以SP2/AP1、SP2/AP2、SP2/AP3和SP2/AP4为引物对, PCR扩增目的片段。取5 μL PCR产

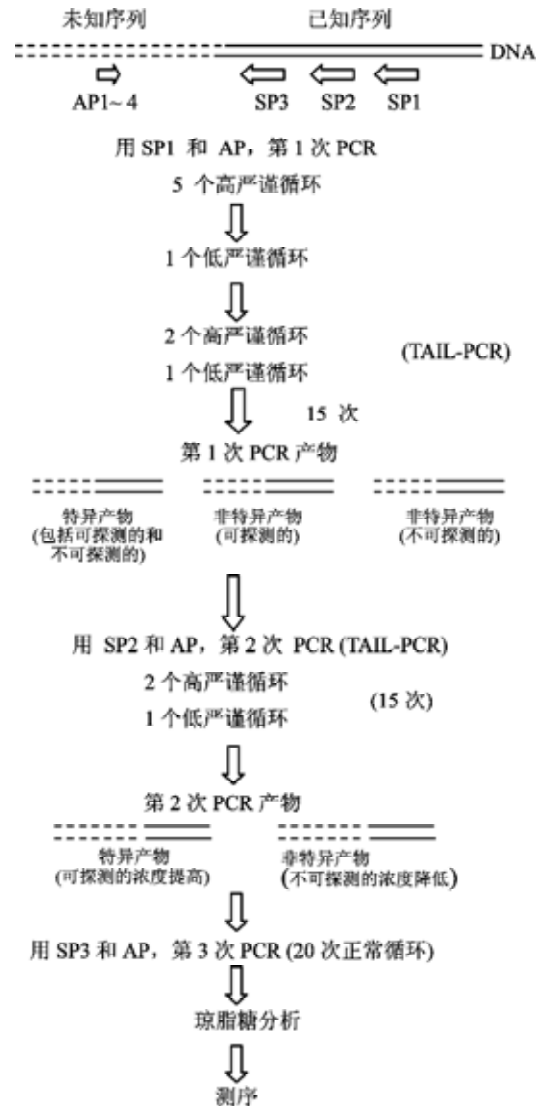


图1 TAIL-PCR的原理(Liu 和 Whittier 1995)

物进行电泳检测, 其余作为第3次PCR的模板。

第3次PCR: 以第2次PCR产物为模板, 分别以SP3/AP1、SP3/AP2、SP3/AP3和SP3/AP4为引物对, PCR扩增目的片段。取5 μL PCR产物进行电泳检测。

3次扩增的反应体系、温度条件和循环数见表1。

5 产物回收和测序

用TaKaRa琼脂糖胶DNA片段回收试剂盒(Code No. D301)回收第3次PCR产物较长片段, 然后取1 μL进行琼脂糖凝胶电泳, 并用SEQR1、SEQR2和SEQR4引物对回收产物测序。

表1 TAIL-PCR的扩增体系、温度条件和循环数

反应	循环体系中物质含量/ μL	温度条件	循环数
初次反应	10 \times LA PCR 缓冲液 II (加 Mg^{2+})	5 93 (1 min), 95 (1 min)	1
	DNTP 混合物(2.5 $\text{mmol}\cdot\text{L}^{-1}$)	8 94 (30 s), 60 (1 min), 72 (2 min)	5
	SP1 (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 94 (30 s), 25 (3 min), 72 (2 min)	1
	AP (1~4) (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 94 (30 s), 60 (1 min), 72 (2 min)	15
	基因组 DNA	1 94 (30 s), 60 (1 min), 72 (2 min)	1
	TaKaRa LA <i>Taq</i> (5 $\text{U}\cdot\text{L}^{-1}$)	0.5 94 (30 s), 44 (1 min), 72 (2 min)	1
	加双蒸水到	50 72 (10 min)	1
二次反应	10 \times LA PCR 缓冲液 II (加 Mg^{2+})	5 94 (30 s), 60 (1 min), 72 (2 min)	1
	DNTP 混合物(2.5 $\text{mmol}\cdot\text{L}^{-1}$)	8 94 (30 s), 60 (1 min), 72 (2 min)	1
	SP2 (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 94 (30 s), 44 (1 min), 72 (1 min)	1
	AP (1~4) (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 72 (10 min)	1
	初次 PCR 产物	1	1
	TaKaRa LA <i>Taq</i> (5 $\text{U}\cdot\text{L}^{-1}$)	0.5	
	加双蒸水到	50	
三次反应	10 \times LA PCR 缓冲液 II (加 Mg^{2+})	5 94 (30 s), 60 (1 min), 72 (2 min)	12
	DNTP 混合物(2.5 $\text{mmol}\cdot\text{L}^{-1}$)	8 94 (30 s), 60 (1 min), 72 (2 min)	1
	SP3 (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 94 (30 s), 44 (1 min), 72 (1 min)	1
	AP (1~4) (20 $\text{pmol}\cdot\text{L}^{-1}$)	1 72 (10 min)	1
	二次 PCR 产物	1	1
	TaKaRa LA <i>Taq</i> (5 $\text{U}\cdot\text{L}^{-1}$)	0.5	
	加双蒸水到	50	

6 侧翼序列验证克隆

用 TaKaRa LA *Taq*TM DNA 聚合酶, 以小麦基因组 DNA 为模板, 以 SEQR5/SEQR3 为引物, PCR 扩增目的基因, 用琼脂糖胶 DNA 片段回收试剂盒回收 PCR 产物片段, 取 1 μL 进行琼脂糖凝胶电泳。

用 TaKaRa DNA Ligation Kit (Code No. D6022) 中的溶液 I, 将回收产物与 pMD18-T 样品载体连接后, 热转化至大肠杆菌 JM109 感受态细胞 (Code No. D9052), 涂布平板, 37 $^{\circ}\text{C}$ 过夜培养。挑选阳性菌落扩大培养, 提取质粒后取 1 μL 进行琼脂糖凝胶电泳。用 BacBEST 引物 RV-M、M13-47 和 SEQR1、SEQR3 引物对质粒进行测序。

实验结果

1 三次 PCR 产物的电泳分析

图 2 为 3 次 PCR 产物电泳图。结果表明, 不同 PCR 所得到的产物在大小和丰度方面存在明显差异, 片段最小为 100 bp, 最大为 10 000 bp。第 1 次 PCR 产物中 SP1 分别与 AP1 和 AP2 组成的

引物对未检测到有扩增产物(泳道 1、2), SP1 分别与 AP3 和 AP4 组成的引物对分别检测到有 2 条扩增产物(泳道 3、4), 大小在 750~2 000 bp 之间, 其中 SP1 与 AP3 组成的引物对检测到扩增产物丰度最大(图 2-a)。第 2 次 PCR 产物中 SP2 与 AP1 组成的引物对未检测到有扩增产物(泳道 1), SP2 分别与 AP2、AP3 和 AP4 组成的引物对分别检测到有 5 条、6 条、7 条扩增产物(泳道 2、3、4), 大小均在 200 bp 以上, 其中 SP2 与 AP3 组成的引物对检测到扩增产物中大约 2 000 bp 的片段丰度最大(图 2-b)。第 3 次 PCR 产物中 SP3 分别与 AP1、AP2、AP3 和 AP4 组成的引物对分别检测到有 1、5、5 和 6 条扩增产物(泳道 1、2、3 和 4), 其中以 SP3 与 AP3 为引物检测到扩增产物中有大约 5 000、2 000、1 100、750 和 600 bp 的片段(图 2-c)。

2 目的片段的回收和测序
切胶回收图 2-c 泳道 3 中最富集的 PCR 产物, 其大小为 5 000 bp 左右的特异片段, 重新进行琼脂糖凝胶电泳(图 3-a)。以上述片断为模板, 以

SEQR5/SP3 为引物对侧翼序列进行 PCR 验证, 扩增产物片段大小接近 2 000 bp (图 3-b), 说明该片段为目的片段。

将目的片段装入大小为 2 692 bp 的克隆载体 pMD18-T 中(图 3-c), 用测序引物 SEQR1、SEQR2 和 SEQR4 对回收产物进行测序。结果表明, 回收的目的片段全长 1 748 bp。根据启动子的结构

特征, 对 CAAT-box 和 TATA-box 等启动子元件进行分析表明, 它们分别位于基因转录起始位点 CAT-box 上游 -163 bp 和 -27 bp 处, ATG 位于 CAT-box 下游 94 bp 处, 5'UTR (非编码区) 序列全长 93 bp, 说明该序列确为小麦 *GLP3* 启动子序列。测序结果已登录于 GenBank, 登录号为 AY864922。其结构示意图见图 4。

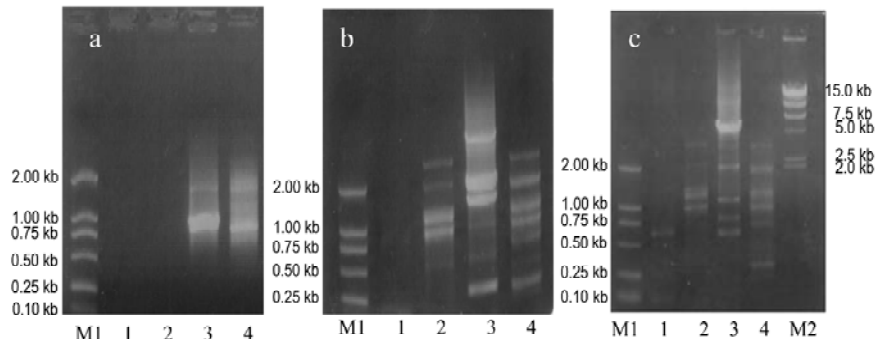


图2 TAIL-PCR 产物的琼脂糖凝胶电泳分析

a、b、c 分别代表第 1、2、3 次 PCR 产物电泳结果。M1 : DNA 分子量标准 DL2000; M2 : DNA 分子量标准 λ -HindIII ; 1~4 分别代表不同泳道中特异引物与随意引物 AP1~4 组合的 PCR 产物。

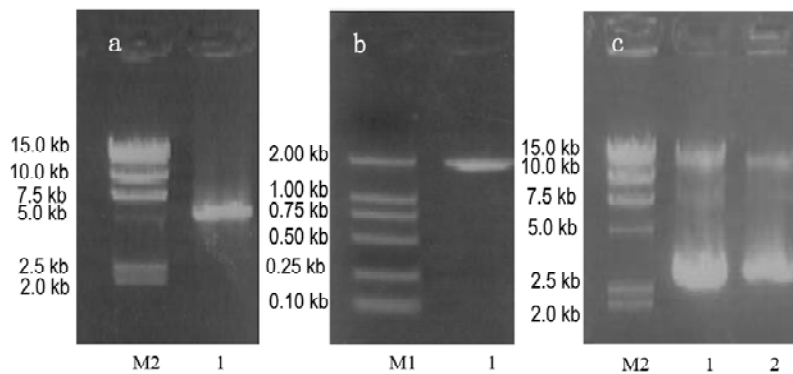


图3 目的片段回收(a, 泳道 1)、侧翼序列验证(b, 泳道 1)与克隆(c, 泳道 1 和 2)
M1 : DNA 分子量标准 DL2000 ; M2 : DNA 分子量标准 λ -HindIII。

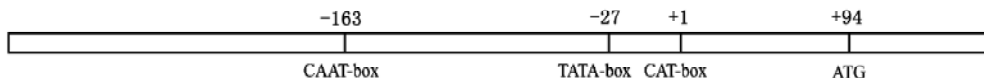


图4 *GLP3* 启动子序列示意

讨 论

虽然用随机引物进行染色体步行的技术提出较早, 但由于无法有效地控制由随机引物引发的

非特异产物的产生, 这种方法一直未能得到广泛应用。Liu 和 Whittier (1995) 提出的 TAIL-PCR 具有过程简单、快速省时、高效特异和直接测序等优点, 于是这个问题便得到了解决。

本文采用的方法是在 Liu 和 Whittier (1995) 方法的基础上作了如下改进:(1)省去第 1 次 TAIL-PCR 前的 10 个低严谨度的 PCR 循环,可以节省时间;(2) 3 次 PCR 循环中的变性时间由 5 s 延长至 30 s,这对 DNA 彻底解链有利;(3) 3 次 PCR 循环中的延伸时间由 5 min 延长至 10 min,可以获得较长的侧翼序列;(4)第 3 次循环中增加加热不对称交错 PCR,便于特异目的片段的富集,这也是与 Liu 和 Whittier (1995) 方法的最大不同之处。只须经过 1 次反应(约 12 h)就可克隆到 *GLP3* 基因启动子的全长序列(1 748 bp)。

McCubbin 等(1987)最初在小麦中鉴定出 *Germin* 基因, Rahman 等(1988)相继分离出全长 *Germin* 的 cDNA 序列;之后, Dratewka 等(1989)又进行 cDNA 的测序。Lane 等(1991)用全长的 *Germin* cDNA 探针在小麦 DNA 文库中同时克隆到小麦 *Germin* 的 2 个基因,二者 mRNA 的同源性为 87%。并观察到 *Germin* 在不同生长发育阶段(如花期诱导、胚胎发生、次生木质部的形成)或在生物或非生物胁迫条件下都会表达,并表现出生物学功能(Fry 等 1998; Caliskan 等 2004)。Chen 等(2006)报道 *Germin* 活性与小麦愈伤组织的形成和分化有关,其生理作用是多方面的,但对其启动子的研究较少。本文用 TAIL-PCR 的方法成功克隆到 *GLP3* 基因启动子的全长序列(1 748 bp),对相关问题的研究可能有一定的参考价值。

参考文献

- Antal Z, Rasclé C, Fevre M, Bruel C (2004). Single oligonucleotide nested PCR: a rapid method for the isolation of genes and their flanking regions from expressed sequence tags. *Curt Genet*, 46 (4): 240~246
- Benkel BF, Fong Y (1996). Long range inverse PCR (LR-IPCR): Extending the useful range of inverse PCR. *Genet Anal*, 13 (5): 123~127
- Caliskan M, Turet M, Cuming AC (2004). Formation of wheat (*Triticum aestivum* L.) embryogenic callus involves peroxide-generating germin-like oxalate oxidase. *Planta*, 219: 132~140
- Carter C, Thornburg RW (1999). Germin-like proteins: structure, phylogeny, and function. *J Plant Biol*, 42: 97~108
- Chen JY, Yue RQ, Xu HX, Chen XJ, Cui DQ (2006). Study on plant regeneration of wheat mature embryos under endosperm supported culture. *Agr Sci Chin*, 5 (8): 572~578
- Dratewka-Kos E, Rahman S, Grzelczak ZF, Kennedy TD, Murray RK, Lane BG (1989). Polypeptide structure of germin as deduced from cDNA sequencing. *J Biol Chem*, 264: 4896~4900
- Fry SC (1998). Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem J*, 332: 507~515
- Jones DH, Winistorfer SC (1992). Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucleic Acids Res*, 20 (3): 595~600
- Jones DH, Winistorfer SC (1997). Amplification of 4-9 kb human genomic DNA flanking a known site using a panhandle PCR variant. *Biotechniques*, 23 (1): 132~138
- Kim HJ, Triplett BA (2004). Cotton fiber germin-like protein. I. Molecular cloning and gene expression. *Planta*, 218: 516~524
- Lagerstrom M, Parik J, Malmgren H, Stewart J, Petersson U, Landegren U (1991). Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Methods App1*, 1: 111~119
- Lane BG, Bernierp F, Dratewka-Kos E, Shafai R, Kennedy TD, Pyne C, Munro J, Vaughan T, Walters D, Altomare F (1991). Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain *Physarum* spherulins. *J Biol Chem*, 266 (5): 10461~10469
- Liu YG, Whittier RF (1995). Thermal asymmetric interlace PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, 25 (3): 674~681
- Mazars GR, Moyret C, Jeanteur P, Theillet CG (1991). Direct sequencing by thermal asymmetric PCR. *Nucleic Acids Res*, 19: 4783
- McCubbin WC, Cyril MK, Kennedy TD, Lane BG (1987). Germin: physicochemical properties of the glycoprotein which signals the onset of growth in the germinating wheat embryo. *Biochem Cell Biol*, 65: 1039~1048
- Ochman H, Gerber AS, Hantl DL (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics*, 120: 621~623
- Parker JD, Rabinovitch PS, Burmer GC (1991). Targeted gene walking polymerase chain reaction. *Nucleic Acids Res*, 19: 3055~3060
- Qin GJ, Kang DM, Dong YY, Shen YP, Zhang L, Deng XH, Zhang Y, Li S, Chen N, Niu WR (2003). Obtaining and analysis of flanking sequences from T-DNA transformants of *Arabidopsis*. *Plant Sci*, 165 (5): 941~949
- Rahman S, Grzelczak Z, Kennedy T, Lane BG (1988). Germin. Molecular cloning of cDNA that selects germin mRNA from bulk wheat mRNA. *Biochem Cell Biol*, 66: 100~106
- Rosenthal A, Jones DSC (1990). Genomic walking and sequencing by oligocassette mediated polymerase chain reaction. *Nucleic Acids Res*, 18: 3095~3096
- Sarkar G, Turner RT, Bolander ME (1993). Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods App1*, 2: 318~322
- Woo EJ, Dunwell JM, Goodenough PW, Marvier AC, Pickersgil RW (2000). Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat Struct Biol*, 7: 1036~1040
- Yan YX, An CC, Li L, Gu JY, Tan GH, Chen ZL (2003). T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. *Nucleic Acid Res*, 31 (12): e68