

Phân tích trên quy mô hệ gene họ NIN-like protein (NLP) ở cây ngô (*Zea mays* L.) bằng phương pháp tin sinh học

• **Cao Phi Bằng**

Trường Đại học Hùng Vương, Phú Thọ

(Bài nhận ngày 22 tháng 08 năm 2016, nhận đăng ngày 04 tháng 05 năm 2017)

TÓM TẮT

Các *NIN-like protein (NLP)* thuộc họ tác nhân điều hòa phiên mã *RWP-RK* và mang các đặc điểm giống với *NIN (Nodule INception)*. Các *NLP* điều hòa sự biểu hiện các gene liên quan đến tín hiệu nitrate ở thực vật. Trong công trình này, chúng tôi phân tích họ gene *NLP* trên quy mô hệ gene của cây ngô (*Zea mays* L.) thông qua cách tiếp cận tin sinh học. Trong hệ gene của cây ngô có tổng số chín gene mã hóa *NLP*. Các gene này có ba hoặc bốn intron, dài từ 2855 tới 8092 nucleotide. Các trình tự protein suy diễn có chiều dài từ 786 tới 945 amino acid. Giá trị *pI* của hầu

Từ khóa: *NIN-like protein (NLP)*, biểu hiện gene, cấu trúc, tiến hóa, cây ngô

MỞ ĐẦU

Tác nhân điều hòa phiên mã *NIN (Nodule INception)* được phát hiện lần đầu ở cây đậu đại năm 1999, giữ vai trò quan trọng trong sự hình thành nốt sần [6]. Phân tử protein *NIN* có vùng bảo thủ gắn DNA đặc trưng là vùng *RWP-RK* [6]. Các protein gần giống *NIN (NIN-Like Protein, NLP)* có vùng *C-terminal* và vùng *N-terminal* tương đồng với *NIN* [7]. Tuy nhiên, vai trò của các *NLP* này chủ yếu là điều hòa các gene cảm ứng bởi tín hiệu nitrate [5]. Sự tiến hóa của các *NLP* của các cây *A. thaliana*, đậu đại và lúa được báo cáo bởi Schauser *et al.* (2005) [7]. Trong báo cáo này, các cây *A. thaliana*, đậu đại và lúa lần lượt có chín, năm và ba gene thuộc về họ *NLP*.

Cây ngô được biết là một trong những cây lương thực quan trọng nhất trên thế giới. Hệ gene của loài cây này được hoàn thành giải trình tự

hết các protein đều nhỏ hơn 7. Các protein *NLP* của cây ngô có các vùng bảo thủ cao phía đầu *N-terminal* và đầu *C-terminal* trong đó có vùng *RWP-RK* và *PB1* đặc trưng. Căn cứ vào cây phả hệ có thể nhận thấy ba sự kiện nhân gene trên quy mô hệ gene mới diễn ra sau thời điểm phát sinh loài ở cây ngô. Tất cả các gene *NLP* biểu hiện ở tất cả các mô của cây ngô ở các giai đoạn phát triển khác nhau, từ giai đoạn hạt nảy mầm tới hạt chín đều được nghiên cứu. Các gene *ZmNLP5*, *ZmNLP6* và *ZmNLP7* là những gene biểu hiện yếu hơn so với các gene còn lại ở phần lớn các mô.

năm 2009 [8], nhờ đó chúng ta cũng có thể phân tích họ gene *NLP* hay bất kỳ họ gene nào khác trên quy mô hệ gene của cây ngô. Chúng tôi cảm ơn các nhà khoa học đã giải trình tự hệ gene cũng như hệ mã phiên của loài cây lương thực quan trọng này, mở đường và thúc đẩy nhiều nghiên cứu ở mức phân tử cũng như việc ứng dụng kết quả nghiên cứu vào thực tiễn sản xuất. Trong nghiên cứu này, chúng tôi hướng tới việc sử dụng phương pháp nghiên cứu tin sinh học để phân tích các đặc điểm của họ gene *NLP* ở cây ngô. Những phân tích trên quy mô hệ gene sẽ bổ sung nhiều thông tin khoa học về họ *NLP* vốn còn chưa biết ở loài cây này.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Cơ sở dữ liệu về trình tự hệ gene của cây lúa và cây ngô

Trong nghiên cứu này chúng tôi sử dụng hệ gene đã được giải trình tự của cây ngô [8].

Xác định và phân tích các gene thuộc họ *NLP* ở cây ngô

Chúng tôi sử dụng gene *NIN* của cây đậu đại (*Lotus japonica*) [6] làm khuôn dò để tìm kiếm các gene tương đồng trong hệ gene của cây ngô. Các protein suy diễn được kiểm tra vùng bảo thủ với Pfam [1]. Các đặc điểm của các protein nghiên cứu được khảo sát nhờ các công cụ của ExPASy [2]. Cấu trúc exon/intron được xây dựng nhờ GSDS (Gene Structure Display Server) 2.0 [3].

Xây dựng cây phả hệ

Cây phả hệ được xây dựng từ các protein *NLP* của cây ngô (Pv), cây đậu đại (Lj), cây *A. thaliana* (At) và cây lúa (Os) nhờ sử dụng phần mềm MEGA5 [11], sau khi chúng được sắp dãy bằng MAFFT (Multiple Alignment using Fast Fourier Transform) [4], với các tham biến: thuật toán Maximum Likelihood, mô hình Jones-Taylor-Thornton (JTT), phương pháp Bootstrap với 1000 lần lặp lại.

Phân tích sự biểu hiện gene

Cơ sở dữ liệu hệ mã phiên của cây ngô được sử dụng để phân tích sự biểu hiện của các gene *NLP* được lấy từ dữ liệu microarray của Sekhon *et al.* (2010) [9]

KẾT QUẢ VÀ THẢO LUẬN

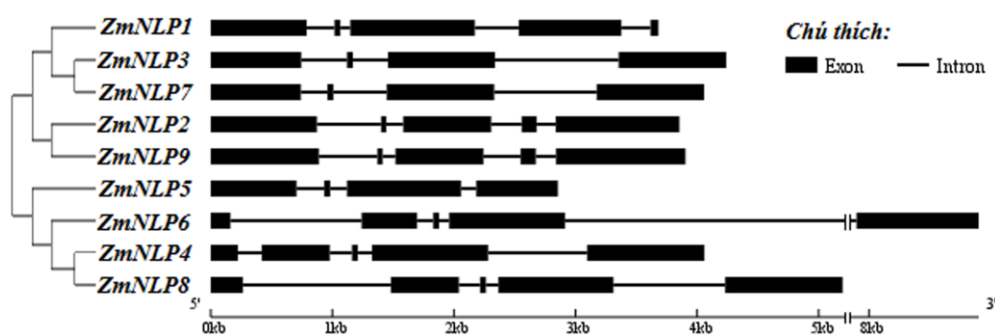
Các *NLP* ở cây ngô

Trên quy mô hệ gene, chúng tôi phát hiện cây ngô có tổng số chín gene mã hóa *NLP* khi tìm kiếm các gene tương đồng với khuôn dò là gene *NIN* của cây đậu đại. Để kiểm tra sự có mặt các vùng bảo thủ trong trình tự protein suy diễn, chúng tôi sử dụng Pfam, server chứa dữ liệu về các họ protein có vùng bảo thủ được xác định [1]. Các protein suy diễn có mang vùng bảo thủ RWP-RK đặc trưng cho các *NLP*. Khi so sánh với một số cây khác, kích thước họ gene *NLP* của cây ngô lớn hơn đậu đại (5 gene) và bằng với cây *A. thaliana* (9 gene) [7].

Các gene mã hóa *NLP* ở cây ngô có chiều dài khác nhau, trong khoảng từ 2855 nucleotide (gene *ZmNLP5*) tới 8092 nucleotide (gene *ZmNLP6*) (Bảng 1). Cấu trúc exon/intron của các gene mã hóa *NLP* của cây ngô được xây dựng nhờ GSDS 2.0 [3] (Hình 1). Các gene *NLP* của cây ngô mã hóa không liên tục, có ba (*ZmLNP3*, *ZmLNP5* và *ZmLNP7*) hoặc bốn intron (*ZmNLP1*, *ZmLNP2*, *ZmLNP4*, *ZmLNP6*, *ZmLNP8* và *ZmLNP9*).

Bảng 1. Các gene thuộc họ *NLP* của cây ngô và đặc điểm của chúng

Gene	Tên locus	Kích thước gene (bp)	Chiều dài protein (aa)	Khối lượng phân tử protein (kDa)	pI	Nhiễm sắc thể	Số lượng intron
<i>ZmLNP1</i>	GRMZM2G109509	3681	922	102,7	5,6	1	4
<i>ZmLNP2</i>	GRMZM2G031398	3854	925	100,6	6,0	2	4
<i>ZmLNP3</i>	GRMZM2G048582	4239	851	93,3	6,5	2	3
<i>ZmLNP4</i>	GRMZM2G475305	4060	916	101,0	6,0	3	4
<i>ZmLNP5</i>	GRMZM2G042278	2855	786	82,9	5,9	3	3
<i>ZmLNP6</i>	GRMZM2G176655	8902	873	97,6	7,7	6	4
<i>ZmLNP7</i>	GRMZM2G053298	4058	851	93,4	6,3	7	3
<i>ZmLNP8</i>	GRMZM2G375675	5196	927	102,1	5,6	8	4
<i>ZmLNP9</i>	GRMZM2G105004	3904	945	102,8	6,1	10	4



Hình 1. Cấu trúc exon/intron của các NLP của cây ngô

Các protein suy diễn NLP của cây ngô có chiều dài từ 786 tới 945 amino acid, khối lượng phân tử của chúng ở trong khoảng từ 82,9 tới 102,8 kDa. Giá trị pI của các protein này nhỏ hơn 7 ngoại trừ ZmNLP6, protein này có pI bằng 7,7 (Bảng 1). Các đặc điểm lí hóa của các NLP lần đầu tiên được mô tả trong báo cáo này.

Các vùng bảo thủ của các protein NLP của cây ngô được thể hiện ở Hình S1. Sáu vùng bảo thủ của các protein NLP của cây ngô tương tự như các vùng bảo thủ đã được mô tả ở cây đậu đại cũng như cây *A. thaliana* [7, 10]. Các vùng bảo thủ này được đánh dấu bằng nền xám. Vùng bảo thủ ở đầu N-terminal của các NLP của cây ngô giống nhau, không bị biến đổi nhiều trong quá trình tiến hóa. Hiện tượng này giống ở cây *A. thaliana* nhưng khác với ở cây đậu đại [7], loài cây thuộc họ Đậu có khả năng cố định đạm nhờ các vi sinh vật cộng sinh trong nốt sần. Ở loài này, có một đoạn trong vùng bảo thủ phía đầu N-terminal bị biến đổi nên không còn chức năng điều hòa các gene cảm ứng với nitrate nhưng tham gia cảm ứng sự hình thành nốt sần [7, 10]. Hai vùng bảo thủ cuối cùng ở đầu C-terminal lần lượt là vùng RWP-RK và vùng PB1, vùng RWP-RK có chức năng gắn với deoxyribonucleic acid (DNA), vùng PB1 giữ chức năng tương tác protein - protein [7, 10].

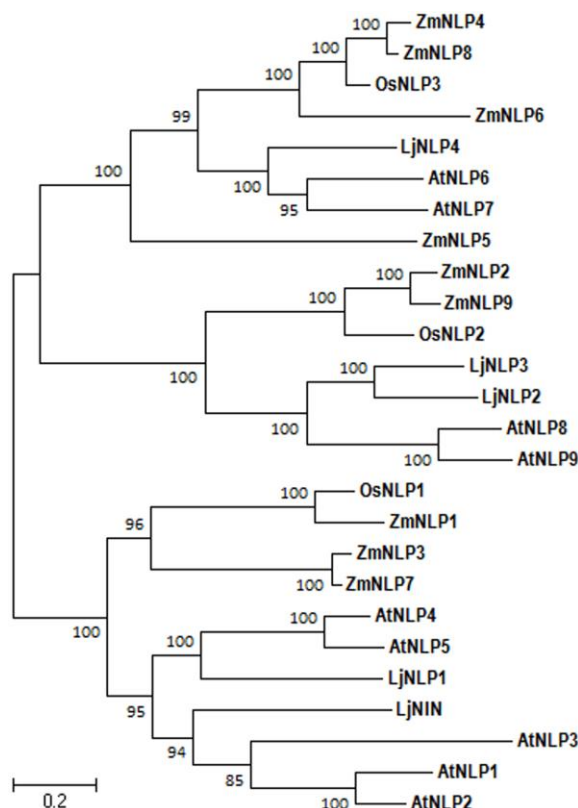
Cây phả hệ và tiến hóa các NLP ở cây ngô

Cây phả hệ từ các protein NLP của *A. thaliana*, cây đậu đại, cây lúa [7] và cây ngô được xây dựng trong nghiên cứu này (Hình 2).

Trên cây phả hệ có ba nhánh, mỗi nhánh đều có ít nhất một đại diện của các loài. Kết quả này xác định giả thuyết có ít nhất ba gene NLP đã tồn tại ở cây tổ tiên chung của thực vật một lá mầm và thực vật hai lá mầm được nêu bởi Schauser *et al.* (2005) [7]. Thực vậy, cây phả hệ được xây dựng từ các protein NLP ở cây *A. thaliana* cũng như cây lúa và cây đậu đại trong nghiên cứu của Schauser *et al.* (2005) có ba nhánh. Mỗi nhánh đều có các đại diện NLP của cây lúa và cây *A. thaliana*. Riêng các đại diện NLP của cây đậu đại chỉ phân bố vào hai trong ba nhánh của cây phả hệ. Do trong nghiên cứu của mình, Schauser *et al.* (2005) chỉ xác định được ba gene NLP từ ngân hàng DNA bổ sung (cDNA). Nhờ sự có sẵn hệ gene cây đậu đại, chúng tôi đã xác định thêm được hai gene khác (*LjNLP3* và *LiNLP4*), trong đó gene *LjNLP4* phân bố vào nhánh thứ ba của cây phả hệ. Trên mỗi nhánh của cây phả hệ được xây dựng trong nghiên cứu này đều có các gene NLP của cây ngô, lúa (cây một lá mầm) và các cây *A. thaliana*, đậu đại (cây hai lá mầm), có thể thấy rằng có ít nhất ba gene NLP tổ tiên chung của cây một lá mầm và hai lá mầm và chúng đã tiến hóa khác nhau sau thời điểm phát sinh loài ở hai lớp thực vật này. Ở nhánh có chứa gene *NIN* của các cây họ Đậu, cây ngô có ba gene. Số gene của cây lúa, cây *Arabidopsis* và cây đậu đại lần lượt là một gene, năm gene và hai gene. Ở nhánh giữa, ngoài cây lúa có một gene, các loài còn lại có số lượng gene bằng nhau (cùng có hai gene). Cuối cùng, ở nhánh còn lại, cây ngô có nhiều

gene hơn so với các cây khác. Vị trí của các gene *NLP* của cây ngô trên cây phả hệ cũng cho phép xác định ít nhất ba sự kiện nhân gene trên quy mô hệ gene sau thời điểm biệt hóa loài giữa cây lúa

và cây ngô (hình thành các cặp gene *ZmNLP2* và *ZmLNP9*, *ZmNLP3* và *ZmNLP7*, *ZmNLP4* và *ZmLNP8*).



Hình 2. Cây phả hệ được xây dựng từ các protein NLP của cây ngô (Pv), cây đậu đại (Lj), cây *A. thaliana* (At) và cây lúa (Os) với các tham biến: thuật toán Maximum Likelihood, mô hình Jones-Taylor-Thornton (JTT), phương pháp Bootstrap với 1000 lần lặp lại, giá trị bootstrap được thể hiện trên mỗi nhánh, tỷ lệ xích là số amino acid thay thế trên một vị trí

Sự biểu hiện của các gene *NLP* trong các mô của cây ngô

Chúng tôi phân tích sự biểu hiện của các gene *NLP* của cây ngô trong hệ mã phiên ở cây ngô non, ở cây giai đoạn ra hoa và tạo hạt được xây dựng bởi Sekhnon *et al* (2011) [9]. Trong đó các tác giả đã phân tích một cách tổng quan hệ mã phiên, đưa ra các thông tin tổng quan về hệ mã phiên ở các mô khác nhau cũng như ở nhiều giai đoạn phát triển. Bên cạnh đó, các tác giả đã bước đầu tập trung phân tích sự biểu hiện các gene liên quan đến quá trình sinh tổng hợp lignin.

Sự biểu hiện của các gene khác, trong đó có các *NLP*, chưa được các tác giả phân tích một cách chi tiết. Hệ mã phiên của cây ngô do Sekhnon *et al.* (2011) là một tài liệu quý, cho phép khám phá cũng như phân tích chức năng của các gene ở cây ngô. Với mục đích bước đầu tìm hiểu chức năng của các gene *NLP* ở cây ngô, chúng tôi đã khai thác hệ mã phiên này để tìm hiểu sự biểu hiện các gene được nghiên cứu.

Bảng 2. Sự biểu hiện của các gene *NLP* trong các mô của cây ngô non microarrays (đơn vị tính: số bản mã phiên)

Gene	Hạt nảy mầm 24 h	Cây mầm	Rễ sơ cấp cây mầm	Rễ sơ cấp cây 1 lá	Rễ sơ cấp cây 3 lá	Chồi ngọn cây 1 lá	Chồi ngọn cây 3 lá	Chồi ngọn cây 4 lá	Đỉnh sinh trưởng cây 5 lá	Giống thân thứ nhất cây 5 lá	Lá non cao nhất cây 5 lá	Lá đầu tiên và bẹ lá cây 5 lá	Đuôi lá thứ hai cây 5 lá	Gốc lá thứ hai cây 5 lá
<i>ZmLNP1</i>	5510	1970	1361	4620	5248	2989	3024	2585	2653	2843	8066	9852	13017	1728
<i>ZmLNP2</i>	8772	2296	2915	4222	9074	2742	2455	3055	1800	2487	4799	3716	7006	5188
<i>ZmLNP3</i>	3415	6204	7635	8857	16073	8697	10915	9189	8567	10662	8950	12136	12651	7980
<i>ZmLNP4</i>	3896	3457	5347	5513	10449	5751	4536	7438	4047	4727	7916	9585	8532	7306
<i>ZmLNP5</i>	594	748	558	443	252	396	343	454	463	426	268	810	435	395
<i>ZmLNP6</i>	145	534	426	121	891	658	687	718	504	450	81	126	116	114
<i>ZmLNP7</i>	1132	1156	2665	1106	6889	2081	2887	1490	1992	1985	1196	3637	624	863
<i>ZmLNP8</i>	4576	4234	6168	11097	12813	7464	6296	9180	7920	7841	19914	25919	21193	8039
<i>ZmLNP9</i>	20201	2257	1787	3625	4565	2182	1621	3466	1361	1658	4141	7466	7615	2783

Ở hạt nảy mầm và cây ngô non, tất cả các *NLP* đều biểu hiện nhưng với cường độ khác nhau tùy gene và theo các mô khác nhau (Bảng 2). Gene biểu hiện yếu nhất ở hạt nảy mầm và cây mầm là *ZmLNP5*. Gene *ZmLNP6* cũng biểu hiện yếu. Trong khi đó gene biểu hiện mạnh nhất ở hạt nảy mầm là *ZmLNP9* và gene biểu hiện mạnh nhất ở cây mầm là *ZmLNP3*. Gene này cũng là gene biểu hiện mạnh nhất ở rễ sơ cấp. Căn cứ vào động thái biểu hiện của các gene ở rễ sơ cấp theo thời gian phát triển, có thể xếp các gene *NLP* của cây ngô thành ba nhóm, nhóm

gene biểu hiện giảm dần chỉ có một gene, *ZmLNP5*, nhóm gene biểu hiện giảm từ thời kì cây mầm tới cây một lá sau đó biểu hiện tăng lên ở cây ba lá có hai gene, *ZmLNP6* và *ZmLNP7*. Các gene còn lại biểu hiện tăng dần theo thời kì phát triển. Ở đỉnh sinh trưởng, sự biểu hiện của các gene này phức tạp hơn, tuy nhiên hai gene *ZmLNP5* và *ZmLNP6* vẫn là hai gene biểu hiện yếu hơn cả. Xu hướng này còn tiếp diễn ở các mô của cây năm lá. Gene biểu hiện mạnh nhất ở chồi và lóng thân là *ZmLNP3*, gene biểu hiện mạnh nhất ở các mô còn lại là *ZmLNP8*.

Bảng 3. Sự biểu hiện của các gene *NLP* trong các mô của cây ngô thành thực thu được từ microarray (đơn vị tính: số bản mã phiên)

Gene	Cờ 15	Cờ 18	Bao phấn	Bấp 4DAF	Râu ngô 4DAF	Bấp 18	Lá 13	Lá bao 1	Lá bao 2	Lá bao ngoài
<i>ZmLNP1</i>	2989	3378	11497	4530	3167	3890	17587	4240	6395	8802
<i>ZmLNP2</i>	3746	5728	5808	6184	5326	10413	14890	8056	9457	8878
<i>ZmLNP3</i>	6298	4613	7400	6232	10580	3246	15565	7670	3582	4032
<i>ZmLNP4</i>	3954	3648	2574	6837	3495	3900	5364	4196	3775	3237
<i>ZmLNP5</i>	502	156	275	442	656	1545	146	519	262	511
<i>ZmLNP6</i>	966	89	50	1482	648	144	200	140	421	287
<i>ZmLNP7</i>	923	1267	1830	835	1309	1236	1723	1177	2876	3518
<i>ZmLNP8</i>	6959	7391	11848	9648	6513	11482	16842	8513	7759	11361
<i>ZmLNP9</i>	3509	3716	6091	3229	3047	5526	17528	5064	5200	7271

Chú thích : Cờ 15 = Cờ non (cây 15 lá), Cờ 18 = Cờ thành thực (cây 18 lá), Bấp 18 = Bấp ngô non (cây 18 lá), Bấp 4DAF = Bấp ngô trước thụ phấn (4 ngày sau nở hoa), Râu 4DAF = Râu ngô (4 ngày sau nở hoa), Lá 13 = Lá thứ 13 (chín sớm của quả), Lá bao 1 = Lá bao trong cùng của bắp (cây có hoa nở đầu tiên), Lá bao 2 = Lá bao trong cùng của bắp (chín sớm của quả), Lá bao ngoài = Lá bao ngoài của bắp (chín sớm của quả)

Bảng 4. Sự biểu hiện của các gene *NLP* trong hạt (phôi và nội nhũ) thu được từ microarray (đơn vị tính: số bản mã phiên)

Gene	Phôi 16 DAP	Phôi 18 DAP	Phôi 20 DAP	Phôi 22 DAP	Phôi 24 DAP	Nội nhũ 12 DAP	Nội nhũ 14 DAP	Nội nhũ 16 DAP	Nội nhũ 18 DAP	Nội nhũ 20 DAP	Nội nhũ 22 DAP	Nội nhũ 24 DAP
<i>ZmLNP1</i>	2315	1961	2257	2367	2981	2935	2407	3124	2889	3143	3404	3330
<i>ZmLNP2</i>	4735	4443	5510	6394	8659	4606	4902	6297	5489	5765	5858	5516
<i>ZmLNP3</i>	7542	8598	13735	16111	15968	2306	3179	3457	5160	4338	5130	3250
<i>ZmLNP4</i>	3609	3643	4194	3594	3123	1949	1526	1494	1952	1598	1916	2438
<i>ZmLNP5</i>	1208	854	347	313	560	224	373	1398	465	786	605	501
<i>ZmLNP6</i>	1007	949	1147	1504	1205	442	146	131	160	108	262	173
<i>ZmLNP7</i>	3254	3504	5036	4882	5800	292	253	252	468	235	223	275
<i>ZmLNP8</i>	4996	4706	5101	5265	5058	3804	3038	3385	4491	3730	4553	4822
<i>ZmLNP9</i>	3364	3546	4010	5330	6018	2237	1276	996	1034	577	583	713

Chú thích : DAF = ngày sau nở hoa

Ở cây ngô thành thực, tất cả các gene *NLP* tiếp tục biểu hiện ở tất cả các bộ phận nghiên cứu (Bảng 3). Hai gene *ZmNLP5* và *ZmNLP6* biểu hiện yếu nhất trong số các gene *NLP* ở hầu hết các bộ phận. Gene *ZmNLP3* biểu hiện mạnh nhất ở cờ chưa thành thực nhưng ở cờ giai đoạn giảm phân (cây có 18 lá) gene biểu hiện mạnh nhất là *ZmNLP8*. Gene này cũng biểu hiện mạnh nhất trong số các gene ở bao phấn, bắp ngô non, râu ngô, ở lá bao trong cùng của bắp giai đoạn nở hoa, lá bao ngoài của bắp giai đoạn chín sớm của quả. Khi xét sự biểu hiện của một gene ở các loại mô, có năm trong tổng số chín gene biểu hiện

mạnh nhất ở lá thứ 13 của cây ở giai đoạn chín sớm của quả. Ngoài ra cả chín gene *NLP* của cây ngô đều biểu hiện ở trong phôi, nội nhũ và hạt nguyên vẹn trong suốt thời kì phát triển sau thụ phấn (Bảng 4 và Bảng 5). Tuy nhiên, cường độ biểu hiện của các gene này biến đổi phức tạp, không cùng chiều hoặc ngược chiều của sự phát triển. Các gene *ZmNLP5*, *ZmNLP6* và *ZmNLP7* là những gene biểu hiện yếu nhất ở các mô nghiên cứu. Lần đầu tiên sự biểu hiện của các gene *NLP* ở mức cơ sở trong một số lượng lớn các mô được phân tích một cách chi tiết.

Bảng 5. Sự biểu hiện của các gene *NLP* trong hạt (vỏ và hạt) thu được từ microarray (đơn vị tính: số bản mã phiên)

Gene	Hạt 2 DAP	Hạt 4 DAP	Hạt 6 DAP	Hạt 8 DAP	Hạt 10 DAP	Hạt 12 DAP	Hạt 14 DAP	Hạt 16 DAP	Hạt 18 DAP	Hạt 20 DAP	Hạt 22 DAP	Hạt 24 DAP
<i>ZmLNP1</i>	2718	3682	4507	4719	4763	4022	3665	3139	3126	3442	4098	4088
<i>ZmLNP2</i>	4985	7872	5197	6725	7885	6398	6282	8365	5002	5172	6099	5867
<i>ZmLNP3</i>	7308	7729	6529	6696	3282	2717	2275	4887	4346	6522	8010	8134
<i>ZmLNP4</i>	3566	3764	3947	3388	2886	2489	1394	1979	2460	1981	2354	2256
<i>ZmLNP5</i>	623	641	509	645	591	640	479	1378	382	462	415	425
<i>ZmLNP6</i>	806	687	1229	933	629	457	202	141	190	210	332	264
<i>ZmLNP7</i>	1323	1342	1438	1455	1094	836	423	552	543	911	979	1450
<i>ZmLNP8</i>	5971	5886	7950	5965	6296	6407	3942	4914	3683	4651	5024	5146
<i>ZmLNP9</i>	3288	3661	4217	3717	3258	3450	2298	2883	1361	1799	2254	2670

Chú thích : DAP = ngày sau thụ phấn,

KẾT LUẬN

Trong nghiên cứu này, họ gene *NLP* ở cây ngô đã được phân tích trên quy mô hệ gene nhờ sử dụng các phương pháp tin sinh học. Họ *NLP* của cây ngô gồm có chín thành viên. Các protein suy diễn của chúng có vùng bảo thủ tương đồng cao ở đầu *N*-terminal và mang các vùng bảo thủ đặc trưng cho các *NLP* của thực vật RWP-RK và PB1 ở đầu *C*-terminal. Cây phả hệ xác định có ba *NLP* ở cây tổ tiên chung phát sinh thực vật một lá mầm và hai lá mầm. Ít nhất ba sự kiện nhân gene

NLP sau thời điểm phát sinh loài đã được phát hiện ở cây ngô. Sự biểu hiện của các gene trong họ *NLP* ở các cơ quan sinh dưỡng và cơ quan sinh sản của cây ngô cũng được báo cáo trong công trình này. Những kết quả này góp phần bổ sung các thông tin khoa học về họ *NLP* ở thực vật một lá mầm.

Lời cảm ơn: Công trình này được hoàn thành với sự hỗ trợ kinh phí từ chương trình nghiên cứu khoa học cơ bản của Trường Đại học Hùng Vương, tỉnh Phú Thọ.

Genome-wide analysis of NIN-like protein (NLP) family in maize (*Zea mays* L.) by using bioinformatic methods

• **Cao Phi Bang**

Hung Vuong University, Phu Tho

ABSTRACT

The NIN-like proteins (NLP) belong to RWP-RK transcription factor family and possess the similarity characteristics of NIN (Nodules INception). The NLPs regulate the expression of genes which are involved in nitrate signaling in plants. In this work, we have performed a genome-wide analysis of the NLP gene family in maize (Zea mays L.) through the bioinformatic methods. We identified a total of nine NLP encoding genes in whole genome of maize. The genomic sequences of these genes were from 2855 to 8092 nucleotides in length and contained three or four introns. Their predicted protein sizes ranged in size from 786 to 945 amino acids.

Keywords: NIN-like protein (NLP), gene expression, structure, evolution, maize

The theoretical isoelectric point values of most deduced protein were less than 7. The maize NLP proteins possessed conserved regions of plant NLP at N-terminal as well as at C-terminal including the RWP-RK and PB1 domains. Based on the phylogenic analysis, we detected three current whole-genome gene duplication events which occurred in maize genome after speciation point. All of NLP genes expressed in tissues at different development stages, from germinating seed to maturation seed were examined. The ZmNLP5, ZmNLP6 and ZmNLP7 were weakly expressed in comparison to others genes in most examined tissues.

TÀI LIỆU THAM KHẢO

- [1]. R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, *Nucleic Acids Res.*, 4, 2, D222–D230 (2014).
- [2]. E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, In *The Proteomics Protocols Handbook*, Springer, New Jersey, 571–607 (2005).

- [3]. B. Hu, J. Jin, A.Y. Guo, H. Zhang, J. Luo, G. Gao, GSDS 2.0: an upgraded gene feature visualization server, *Bioinformatics*, 31, 8, 1296–1297 (2015).
- [4]. K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.*, 30, 4, 772–780 (2013).
- [5]. M. Konishi, S. Yanagisawa, *Arabidopsis* NIN-like transcription factors have a central role in nitrate signalling, *Nat. Commun.*, 4, 1617 (2013).
- [6]. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, A plant regulator controlling development of symbiotic root nodules. *Nature*, 402, 6758 191–195 (1999).
- [7]. L. Schauser, W. Wieloch, J. Stougaard, Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*, *J. Mol. Evol.*, 60, 2, 229–237 (2005).
- [8]. P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F. Wei, *et al.*, The B73 maize genome: complexity, diversity, and dynamics, *Science*, 326, 1112–1115 (2009).
- [9]. R.S. Sekhon, H. Lin, K.L. Childs, C.N. Hansey, C.R. Buell, N. de Leon, S.M. Kaeppler, Genome-wide atlas of transcription during maize development, *Plant J.*, 66, 4, 553–563 (2011).
- [10]. W. Suzuki, M. Konishi, S. Yanagisawa, The evolutionary events necessary for the emergence of symbiotic nitrogen fixation in legumes may involve a loss of nitrate responsiveness of the NIN transcription factor, *Plant Signal Behav.*, 8, 10 (2013).
- [11]. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28, 2731–2739 (2011).


```

ZmLNp3 -----MEEGDDP-QLPSSIVARTTSEGAGVLDLLEQLLSGDN--G--WLEVVSRSPNSVAFPPPSA-----FFSA-----DATAT-----
ZmLNp7 -----MEEGDDP-QLPSSIVARTTSEGAGVLDLLEQLLSGDN--G--WLEVVSRSPNSVAFPPPSA-----FFSA-----DATAT-----
ZmLNp1 -----MEOAQT-HGGGLLAGCMVDVAVGELLMELNAP-----GPFSDPFGGAGSPV-----CFSPFL-----DICTST-----
ZmLNp8 -----MDF--DP-6-PASAKHDSG--GAMPFSLTSMPLFSVSASPOL--PASS-SWLTTPSPMLFEDRHLLPLDAPSAEPAVAANVVE-----
ZmLNp4 -----MDL--DP-AYPPSACAD-----AMFDSLSLTLF-----SPQUTPIANSPSMILTPHSPMLFEDRHLLPLDAP-----AVFE-----
ZmLNp6 -----MDGVLQR-----MAATLGC-----S--TPCPRSTFR--PFRPCRRGRRAATRIILPC-----LMAV-----SVGGS-----
ZmLNp5 -----MDGVLQR-----DVIDLLDGHGHGD--D--SLCERRSSSIRLSRWL-----LMV-----SVGGS-----
ZmLNp2 MDVPTPPNAGCNSINISGPMQSFDDPFGAAMTSFDGYSELSSPSVADHIFSLINDPS--S--VQMFALWSSSGSSPRASAVREDSFUTYTPGM-----DVTSLAQRI
ZmLNp9 MDVPTPPNAGCNS--SNIGPMQSFDDPFGAAMTSFDGYSELSSPSVADHIFSLINDPS--A--AQQWAMWSSSGSSPRASAVREDSFUTYTPGM-----DGTSSIAQRI
:
ZmLNp3 -TVTTTAAAPAANA-----SWL--IQTGASTPS-----VREFFSQAL--SIYRDTQS-DGDLVQLWVPVNRDDGKVLITTSQGFPLD-HRSDSLIRFREVSRTK
ZmLNp7 -TV-TTAAAPAANA-----SWW--IQTGASPS-----VEERFSQAL--SIYRDTQS-DGDLVQLWVPVNRDDGKVLITTSQGFPLD-HRSDSLIRFREVSRTK
ZmLNp1 -TTTATPPPPPPGCDNDQDPTPPRGVW--FQPRLEVDAT-----VKERLRALERIASLQTC--PGELLAQVWPVTIGD-RQVLITTCGQFPLD-CRNERLANRYTVMK
ZmLNp8 -VQARSNSHTTGRVQDI--NHRWQPHLSLDEDDGTNNS-----FFKEKLTAAL--RYFKEST--DQHLVQVWVAPVNGD-RYVLITTSQGFPLD-HQSIGLLQYRAVSM
ZmLNp4 -VQPRSGNSHTSSKPAEQI--SHRWQPHLSLDEDDGTNNS-----LFKEKLTHAL--RYFKEST--DQHLVQVWVAPVNGD-RYVLITTSQGFPLD-HQSIGLLQYRAVSM
ZmLNp6 -----GWTR--LPCKTTARAS-----PPEKLTVLAL--MYFKEST--KQHLVQVWVAPVNGD-HYMLTSGQGFPLD-HQSIGLLQYRAVSM
ZmLNp5 -----VCVRAGSGERSAS-----GGG--AQDKAGIDGA-----VKERLARAL--RIYKDAAGLDGALVQVWVAPNGD-RVVLATRGQGFPLDPPPRCHRLQYRTVSLA
ZmLNp2 NSASVFNPTVEDRLGKDS--GLV--PNNGSQGSSIIIPRSVGNILADRLMAL--SLFRKSL--SDGVLAQVWVAPVNGD-RVVLATRGQGFPLD--QDLAGREVSRN
ZmLNp9 NSAAASPTGVDRLGKDS--GLV--LPSNGSQGSSIIIPRSVGNILADRLMAL--SLFRKSL--SDGVLAQVWVAPVNGD-RVVLATRGQGFPLD--QDLAGREVSRN
:
ZmLNp3 YQFFADVRSQ-ESPLGPRVFIGRLPEWSPDIRCTSYETARVDKAYLDVHGTMGLPVFEKRSYSCGLVIELIMTKQKINFTSELNTICSAQVNLARSTVSSIPRKFKST--ASYK
ZmLNp7 YQFSADVRSQ-DSPLGPRVFIGRLPEWSPDIRCTSYETARVDKAYLDVHGTMGLPVFEKRSYSCGLVIELIMTKQKINFTSELNTICSAQVNLARSTVSSIPRKFKST--ASYK
ZmLNp1 YQFSADETAR-TDLAMPGRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp8 YMFSDGENA-GDLGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp4 YMFSDAENV-GDLGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp6 YTFADVDSV-ELGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp5 HAFYVQGAENVREGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp2 FLFSVKEEGP-LHLGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
ZmLNp9 FLFSVKEEGP-LHLGLGPRVFIGRVPEWTFDVRFTSYETARVHQAQYFDIRGSVALPIEPFRSRVCLGVELVMTTEKVTNAEIQNICSAKVELDSSSDVSSDRAKVTD--TSYR
:
ZmLNp3 DALPEILEVLAACILTHKPLAQVWVTCQ-----GKRG--RHSNDENRYCISTIDEACFVNEA-----EMGPHACSEHILLRGEVAG--KALITNQCPLDGGSSSTKLEY
ZmLNp7 DALPEILEVLAACILTHKPLAQVWVTCQ-----GKRG--RHSNDENRYCISTIDEACFVNEA-----EMGPHACSEHILLRGEVAG--KALITNQCPLDGGSSSTKLEY
ZmLNp1 ALIPEIVDLVLTVECHKPLAQVWVTCQ-----AKRG--RHTDEKLYCVSTIDEACTYRDL--NVNGPHACSEHILLRGEVAG--RAFATNCPFSRDTASSKIQY
ZmLNp8 AALVILEILITVVECHKPLAQVWVTCQKYRSLAHGGGLKSLCSFDGSCMEVCMSTSDVAFHVIDA--HMWGFDAVCEHLLRQGVVSG--KAFISHKPCPSKDIRKPKCLAY
ZmLNp4 AALVILEILITVVECHKPLAQVWVTCQKYRSLAHGGGLKSLCSFDGSCMEVCMSTSDVAFHVIDA--HMWGFDAVCEHLLRQGVVSG--KAFISHKPCPSKDIRKPKCLAY
ZmLNp6 NVLMVEILITVVECHKPLAQVWVTCQKYRSLAHGGGLKSLCSFDGSCMEVCMSTSDVAFHVIDA--HMWGFDAVCEHLLRQGVVSG--KAFISHKPCPSKDIRKPKCLAY
ZmLNp5 VAMSEVADILNQVECHKPLAQVWVTCQ-----RCSTDTHEASLTAAAGPFFLTGADPNENLGFHACVCEHLLRSLRSGGLVEEAAAARPLFCADVITYSMDAY
ZmLNp2 YASTEILDVLAALCHAMPLALTWPTNSNGTANGFCVAKR--IRLDSQPKSVLRIRHESACTVND--KMGFLHACADRHLEKQGIAG--RALKSNLFFSDPDREYSIEDY
ZmLNp9 SAPTEILDVLAALCHAMPLALTWPTNSNGTANGFCVAKR--IGDPQPKAALRHESACTVND--KMGFLHACADRHLEKQGIAG--RALKSNLFFSDPDREYSIEDY
:
ZmLNp3 PLSHHAKIFRLKGAVALRCTRGTI-----ADVFLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----GEAALMENEINSPTPEKDKD--
ZmLNp7 PLSHHAKIFRLKGAVALRCTRGTI-----ADVFLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp1 PLSHHAKIFRLKGAVALRCTRGTI-----ADVFLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp8 PLVHTARMGLAGCFAICLQSSYTG--DYVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp4 PLVHTARMGLAGCFAICLQSSYTG--DYVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp6 PLVHTARMGLAGCFAICLQSSYTG--DYVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp5 PLAHARFCELSCCLAVCARLRGGDSMDTGGGAGAGDECVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp2 PLAHARFCELSCCLAVCARLRGGDSMDTGGGAGAGDECVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
ZmLNp9 PLAHARFCELSCCLAVCARLRGGDSMDTGGGAGAGDECVLEFFLPTDCEALKEQKVLDSLSGTMRSVC--QTLRVVTDREME-----DETVLENEINSPTPEKDKD--
:
ZmLNp3 -----ELSPGQ--SADRRG--AS-----WTSLAGTSQ--EPALAALRMHGTFSV--GQGPSLSGV
ZmLNp7 -----ELSPGQ--SADRRG--AS-----WTSLAGTSQ--EPALAALRMHGTFSV--GQGPSLSGV
ZmLNp1 NIDVCSGIGVP--ARTASLETSEEVSWIASLCAQKGVKEMDGLPFGS--KQDEG--FS-----VTAGWHTSPVIEPGSIFSGGPHKVKVCTLRDPSS--
ZmLNp8 DAQF-----DGSRL--ESPEDD--RHGSHKFDNENKVLIDITEGQLLT--DQSDQNGTSVGRN--DQSDQNGTSVGRN--DQSDQNGTSVGRN--DQSDQNGTSVGRN--
ZmLNp4 DAQF-----DGSRL--ESPEDD--RHGSHKFDNENKVLIDITEGQLLT--DQSDQNGTSVGRN--DQSDQNGTSVGRN--DQSDQNGTSVGRN--DQSDQNGTSVGRN--
ZmLNp5 HVQH-----NCGEII--ASDERN--MI-VTHSD--SGKRISETESECP--PNTCKNKLFAKN--GCTSDII--HSDERF--HSDERF--HSDERF--HSDERF--
ZmLNp6 -----NADVDVQLEL-----NCGEII--ASDERN--MI-VTHSD--SGKRISETESECP--PNTCKNKLFAKN--GCTSDII--HSDERF--HSDERF--HSDERF--HSDERF--
ZmLNp2 -----NGSCLPTNSESSEHDDQPTESAL-----HDLSGDK--QGDRE--P--AK-----LIVAMGTTTT--D--TEASQGHDEHGH--GQDPSQ--A
ZmLNp9 -----SOTCLPTNSESSEHDDQPTESAL-----HDLSGDK--QGDRE--P--AK-----LIVAMGTTTT--D--TEASQGHDEHGH--GQDPSQ--A
:
ZmLNp3 QATAGSGKRTKRTKAEKTSVSLQIRQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp7 QATAGSGKRTKRTKAEKTSVSLQIRQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp1 -NLEKVEGKRTKRTKAEKTSVSLQIRQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp8 LHKTNKPPERRKGAETKTSLEVIQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp4 LHKTNKPPERRKGAETKTSLEVIQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp6 LKNNKSPKRRKGAETKTSLEVIQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp5 -----VQKAAK--KGEKTVSLEELQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp2 RSSKRVKAEKTSSEKFNLSLQIRQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
ZmLNp9 QRSKRVKAEKTSSEKFNLSLQIRQYFAGSLKDAASLGVCPFTLKRIRCHQGITWPSRKIKRVHSLKLLQOQIDSVHGAETAFOJLTYLKYD
:
ZmLNp3 TDTIS--SDNLSGVSVPPT--NQNLDTFNKQPHKSSNVPSTSHS--SCSHG--SQSPSCSVGATEHAQ--GIIDL
ZmLNp7 TDTIS--SDNLSGVSVPPT--NQNLDTFNKQPHKSSNVPSTSHS--SCSHG--SQSPSCSVGATEHAQ--GIIDL
ZmLNp1 TKTTWS--EKEGLEPVSVPDRYCEGRFTSHTSGNSLSPSCQS--SNSHGSCSGSQTQHVAP--SESSPSCSGGATKHAQ--GIIDQ
ZmLNp8 TQSKVAELSNLAVGDRDSLQKS--LENDHFDVMAQ--GFDNNNDVQLEADKASHSRSSSGESI--NSRT--SESCQSPANQTVCKPI--ASMPAE
ZmLNp4 TQSKVAELSNLAVGDRDSLQKS--LENDHFDVMAQ--GFDNNNDVQLEADKASHSRSSSGESI--NSRT--SESCQSPANQTVCKPI--ASMPAE
ZmLNp6 SQINVSGLPIPSVQENRSDSSQKLLENDTHRTLVLPQ--SFLANINQIEAEKASNRNSKDCRT--HSGR--SESVPSVSPANMTFV--STILE
ZmLNp5 -----APFF--AAAAASVQAPAVAAAPAPALSPCLS--SAPRV--ASSQASQAPPP--
ZmLNp2 GVTKVEKSSPEFGQFSSPDGWRSSQLHASS--IPTRVGEVQMLAINNENGRNVTAGIANIQHNS--EDAHGPLY--FVGAVNSRGTETGYIDPSTSL--MDSEGGQTVTVN
ZmLNp9 G-RAVETCSKPSKQFSSPDGWRSSQLHASS--IPTRVGEVQMLAINNENGRNVTAGIANIQHNS--EDAHGPLY--FVGAVNSRGTETGYIDPSTSL--MDSEGGQTVTVN
:
ZmLNp3 MKSENPKVDS-SIQT--MTEHTSLVHFSVHEAPI-DLLQVTEKANG--AHSSR-SLSSP--NFKQ--TDANRVRATFGSEKVRFRNLNFCNFGQLKHELAKRISI
ZmLNp7 MKSENPKVDS-SIQT--MTEHTSLVHFSVHEAPI-DLLQVTEKANG--AHSSR-SLSSP--NFKQ--TDANRVRATFGSEKVRFRNLNFCNFGQLKHELAKRISI
ZmLNp1 -----QVAKVE--FMENKQSLTAKASHIEL-QMPEERLVTLP--RSHSQV-LSSQ--KVENITQMSKPSKIKAMIGERCIFRQSGWFKELKEELARGI
ZmLNp8 PQLN--PEENKPE--FQEPQLPLSRMLIEDSGSKDLMLPTS--TSQPTLAPFANLVMK--HSGSVTIKASKEDIVRFRPSCSGSVTLKEVAKRLQM
ZmLNp4 PQLN--PEENKPE--FQEPQLPLSRMLIEDSGSKDLMLPTS--TSQPTLAPFANLVMK--HSGSVTIKASKEDIVRFRPSCSGSVTLKEVAKRLQM
ZmLNp6 PQQMYIQOGFTKS--PQEDLRFHGMFVNGSGSLDQCMHITSVAVNGTIIGHGSLTTAQNSGIIITCPGSLISAQ--NSRIVTVKARYKEDLFRPFCSGSVTLKEVAKRLQM
ZmLNp5 -----PKKA-----QDASVTVKARYKEDLFRPFCSGSVTLKEVAKRLQM
ZmLNp2 SR--TKEQTLFSTGTTDSSSGS--ASSQAFKGNPGRALKVS--SSDALTVKATCNGTVRKFLPAMGWYHLLLEEVAKRFL
ZmLNp9 SP--FQQADVMTVGHDHTEQTLFSTGTTDSSSGS--ASSQAFKGNPGRALKVS--SSDALTVKATCNGTVRKFLPAMGWYHLLLEEVAKRFL
:
ZmLNp3 VDTNPLVLKYLDDESEWVMTCDADLQELHVKYK--ADTQ--TIKISVHLAVSP--ATRVTAGHTGLS--
ZmLNp7 VDTNPLVLKYLDDESEWVMTCDADLQELHVKYK--ADTQ--TIKISVHLAVSP--ATRVTAGHTGLS--
ZmLNp1 AREVTVDLKYLDDESEWVMTCDADLQELHVKYK--ADTQ--TIKISVHLAVSP--ATRVTAGHTGLS--
ZmLNp8 D-VGTFDIKYLDDHEWVKLACNDLLECEISRL--SGSH--VIRLIVSDIAA--HFGSSCGSSG--
ZmLNp4 D-VGTFDIKYLDDHEWVKLACNDLLECEISRL--SGSH--VIRLIVSDIAA--HFGSSCGSSG--
ZmLNp6 D-VGTFDIKYLDDHEWVKLACNDLLECEISRA--SGSH--VIRLIVSDIAA--HFGSSCGSSG--
ZmLNp5 A-PGEFVTVLDDNEWVLLSCDADPQCLDVPALSGASASSGTAQVPLVMHVEVAE--VHGSSCGSSD--
ZmLNp2 A-TAGAPQLYKDDDEWVILANDADLQCEVLDLS--IGSR--SVKLQVRLDPC--LVSSGSGSSCC--LQVEGSA-
ZmLNp9 T-TAGAPQLYKDDDEWVILANDADLQCEVLDLS--IGSR--SVKLQVRLDPC--LVSSGSGSSCC--LQVEGSA-
:

```

Hình S1. Kết quả sắp xếp các protein NLP của cây ngô. Các vùng bảo thủ được tô màu xám. Vùng RWP-RK và vùng PB1 được ghi chú trên hình