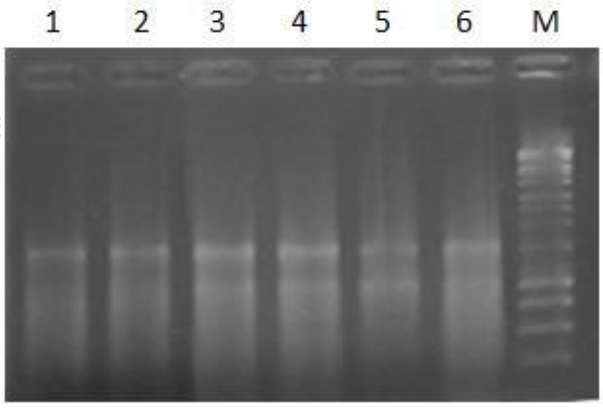
1. GIẢI TRÌNH TỰ HỆ PHIÊN MÃ ĐẶC HIỆU MÔ LÁ VÀ THÂN RỄ SÂM NGỌC LINH 4 NĂM TUỔI

1.1 Kết quả tách chiết, tinh sạch RNA tổng số và xây dựng các thư viện cDNA

RNA tổng số ở các mẫu lá và thân rễ của sâm Ngọc Linh được tách chiết, tinh sạch và đánh giá chất lượng bằng điện di trên gel agarose 8% (

Hình 3.1). Kết quả cho thấy, các mẫu RNA thu được có chất lượng tương đối tốt, có sự xuất hiện của các băng ribosome RNA với kích thước khoảng 1,5- 2 kb. Để tinh sạch một số mẫu RNA, DNase được sử dụng nhằm loại bỏ DNA tổng số.



Hình 3.1 Kiểm tra kết quả tách chiết và tinh sạch RNA tổng số trên gel agarose

M: Thang chuẩn DNA 1 kb; trong đó, số thứ tự tương ứng với các mẫu: 1-3: C4.1-C4.3, 4-6: L4.1-L4.3

Kết quả kiểm tra nồng độ và độ sạch của RNA (Bảng 3.1) cho thấy, các sản phẩm có chỉ số A260/A280 thể hiện sự tinh sạch của mẫu dao động từ 1,90- 2,15 chứng tỏ các mẫu RNA tách chiết được đủ điều kiện để tiến hành những thí nghiệm tiếp theo. Các RNA sau đó được

phân thành những đoạn có kích thước khoảng 450-550 bp sử dụng máy M220 Focused Ultrasonicator, đáp ứng các yêu cầu và được sử dụng để tổng hợp và xây dựng thư viện cDNA phục vụ giải trình tự gen thế hệ mới.

Bảng 3.1 Nồng độ và độ sạch (A260/A280) của các mẫu RNA sâm Ngọc Linh sau tách chiết và tinh sạch

STT	Tên mẫu	Nồng độ (ng/µl)	A260/A280
1	L4.1	1.010,0	1,96
2	L4.2	792,20	2,05
3	L4.3	1.468,3	2,08
4	C4.1	207,10	2,11
5	C4.2	482,40	2,11
6	C4.3	470,80	1,90

1.2 Kết quả giải trình tự hệ phiên mã của mô lá và mô thân rễ sâm Ngọc Linh 4 năm tuổi

Các thư viện được sử dụng cho giải trình tự thế hệ mới Illumina. Để đánh giá chất lượng của các đoạn đọc thô thu được sau khi giải trình tự, phần mềm FastQC được sử dụng để phân tích các chỉ số quan trọng của bộ dữ liệu như số lượng, kích thước của các đoạn đọc, % GC, điểm chất lượng trung bình (Q) của các trình tự theo từng base của các đoạn đọc, mức độ lặp của các đoạn đọc... Điểm chất lượng được biểu diễn dưới dạng biểu đồ hộp, trong đó hộp màu vàng thể hiện 50% số lượng phân bố các điểm chất lượng (gọi là khoảng interquartile), đường màu xanh chỉ giá trị trung bình. Điểm chất lượng trung bình 20 nghĩa là 99% độ chính xác và các đoạn đọc vượt qua điểm 20 được cho là chất lượng tốt.

Điểm chất lượng trung bình từ 28 trở lên được đánh giá là chất lượng đọc rất tốt và đáng tin cậy. Điểm chất lượng trung bình bằng 30 tương đương với độ tin cậy 99,9% [37]. Thông tin thống kê cơ bản của kết quả giải trình tự các thư viện cDNA (Bảng 3.2) cho thấy đã thu được một số lượng lớn các đoạn đọc từ các mẫu mô lá, thân rễ của sâm Ngọc Linh 4 năm tuổi với chất lượng khá cao, có độ tin cậy và đạt yêu cầu cho những phân tích tiếp theo.

Bảng 3.2 Thống kê các bộ dữ liệu thô khi giải trình tự các thư viện cDNA

Mẫu	Tổng số đoạn đọc	Tổng số base đọc	GC (%)	Q20 (%)	Q30 (%)
L4.1	89.888.850	9.078.773.850	43,33	98,89	96,25
L4.2	75.082.240	7.583.306.240	44,53	98,95	96,43
L4.3	66.477.494	6.714.226.894	43,15	98,20	94,39
C4.1	62.126.214	6.274.747.614	44,06	98,89	96,26
C4.2	78.869.048	7.925.373.848	42,70	97,14	92,04
C4.3	63.378.710	6.401.249.710	42,28	98,79	96,04
Tổng	435.822.556	43.977.678.156			

2. PHÂN TÍCH VÀ LẮP RÁP CÁC HỆ PHIÊN MÃ

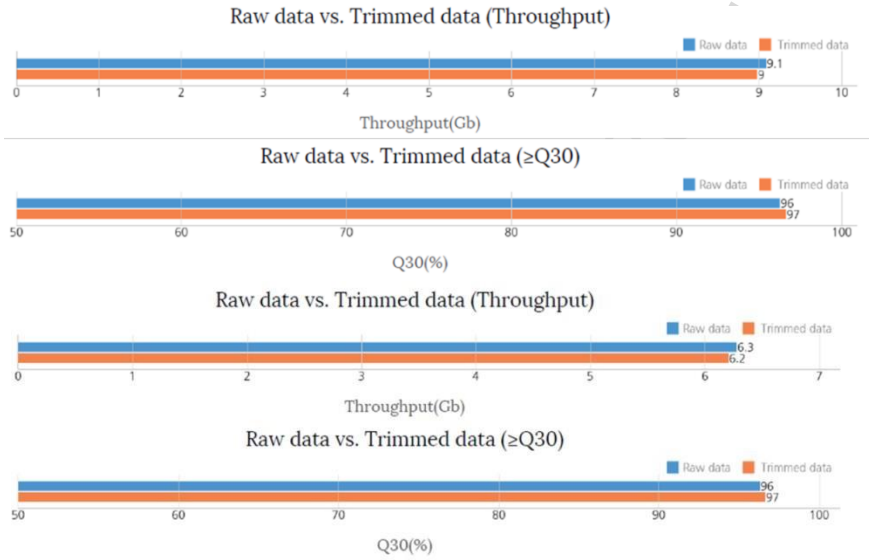
2.1 Kết quả kiểm tra chất lượng các đoạn đọc sau khi trimming

Các bộ dữ liệu thô, sau khi được kiểm tra chất lượng bằng FastQC, tiếp tục được xử lý nhằm loại bỏ trình tự adapter và các base có điểm chất lượng thấp hơn ba từ hai đầu sử dụng Trimmomatic. Theo đó, các đoạn đọc có độ dài ngắn hơn 36 bp sẽ được loại bỏ để tạo ra dữ liệu sau khi trimming (Bảng 3.3). Kết quả, 429.930.834 các đoạn đọc tương ứng với 43.228.319.306 base đã thu được.

Bảng 3.3 Thống kê các bộ dữ liệu thu được sau khi trimming

Mẫu	Tổng số đoạn đọc	Tổng số base đọc	GC (%)	Q20 (%)	Q30 (%)
L4.1	89.141.174	8.975.651.153	43,34	99,12	96,61
L4.2	74.486.684	7.496.686.220	44,53	99,17	96,77
L4.3	65.523.252	6.583.474.260	43,17	98,64	95,01
C4.1	61.627.462	6.203.290.729	44,07	99,12	96,62
C4.2	76.350.630	7.648.769.112	42,74	97,88	93,14
C4.3	62.801.632	6.320.447.832	42,29	99,06	96,45
Tổng	429.930.834	43.228.319.306			

Sau khi trimming, điểm chất lượng trung bình của các trình tự theo từng base của các đoạn đọc cũng được đánh giá lại. So sánh chất lượng đoạn đọc của dữ liệu thô và dữ liệu sau khi trimming cho thấy dữ liệu giải trình tự thô ban đầu có chất lượng khá tốt và bước trimming để loại bỏ những phần không đạt yêu cầu giúp cho kết quả giải trình tự các đoạn đọc được tốt hơn (Hình 3.2). Kết quả lọc chất lượng giúp giảm số lượng đoạn đọc chất lượng thấp thông qua loại bỏ những vùng trình tự kém chất lượng trong các đoạn đọc. Điều này khiến cho số lượng đoạn đọc thu được sau bước lọc giảm nhưng chất lượng của các đoạn đọc lại tăng. Nhìn chung, phân tích điểm chất lượng trung bình của các đoạn đọc chỉ ra kết quả giải trình tự gen thế hệ mới với các mẫu cDNA của mô lá và mô thân rễ sâm Ngọc Linh thu được là tương đối tốt và đủ điều kiện cho lắp ráp và chú giải hệ phiên mã.



Hình 3.2 So sánh dữ liệu thô và dữ liệu sau trimming ở mẫu mô lá (L4.1) và thân rễ (C4.1) của sâm Ngọc Linh 4 năm tuổi

2.2 Kết quả lắp ráp *de novo* các hệ phiên mã

Quá trình phân tích dữ liệu trình tự hệ phiên mã của sâm Ngọc Linh bắt đầu bằng lắp ráp *de novo* các đoạn đọc trình tự đã chọn lọc chất lượng để tạo ra các contig là những đoạn trình tự có kích thước lớn hơn. Ở bước đầu tiên, phần mềm Trinity được sử dụng để chia nhỏ dữ liệu trình tự nhằm phân tích độc lập bằng biểu đồ de Brujin tương ứng với các gen hay locus. Theo đó, các đoạn đọc được lắp ráp gối lên nhau để nối thành những phân đoạn dài hơn mà không chứa các gap hay N. Quá trình lắp ráp *de novo* được thực hiện thông qua 3 module phần mềm riêng biệt của Trinity là Inchworm, Chrysalis và Butterfly. Sau quá trình lắp ráp, phần mềm Trinity sẽ tạo ra một tệp "Trinity.fasta" chứa thông tin về trình tự phiên mã. Các đoạn trình tự thuộc cùng một locus sẽ được phân nhóm thành các cluster dựa trên độ tương đồng về trình tự. Những cluster phiên mã này được tạm coi là các gen.

Bảng 3.4 thể hiện kết quả thống kê của toàn bộ các contig được lắp ráp ban đầu từ các đoạn đọc thu được sau quá trình lọc chất lượng bao gồm số lượng gen, số lượng bản phiên mã (transcript), tỉ lệ GC, chỉ

số N50, kích thước trung bình của contig và tổng số base được lắp ráp. Dữ liệu cho thấy kết quả lắp ráp hệ phiên mã của các mẫu mô lá và thân rễ sâm ở 4 năm tuổi có sự khác nhau về số lượng gen (47.870-70.526), số lượng bản phiên mã (69.638-106.276) và số lượng base lắp ráp (44.043.555-95.505.500). Cụ thể, các mẫu mô lá sâm Ngọc Linh 4 năm tuổi khác nhau về số lượng gen (68.454-81.480), số lượng bản phiên mã (99.609-121.554) và số lượng base lắp ráp (66.828.420-95.505.500); các mẫu thân rễ sâm Ngọc Linh 4 năm tuổi khác nhau về số lượng gen (47.870- 63.268), số lượng bản phiên mã (69.638-100.651) và số lượng base lắp ráp (44.043.555-69.878.264). Kết quả lắp ráp chi tiết các đoạn đọc của các mẫu mô lá và thân rễ sâm Ngọc Linh 4 năm tuổi được trình bày trên Bảng 3.5. Các chỉ số như số lượng gen, số lượng bản phiên mã, tỉ lệ GC, chỉ số N90-N10, kích thước các contig và số lượng base lắp ráp được thống kê cho toàn bộ các contig thu được và cho riêng contig có isoform với kích thước lớn nhất sau lắp ráp. Trong đó, chỉ số N50 thể hiện cho ít nhất 50% các nucleotide được lắp ráp được tìm thấy trong các contig có chiều dài nhỏ nhất và cho isoform dài nhất tương ứng với các giá trị 993, 915, 1.074 của các mẫu mô lá sâm Ngọc Linh 4 năm tuổi; 861, 927, 816 của các mẫu thân rễ sâm Ngọc Linh 4 năm tuổi.

Bảng 3.4 Kết quả thống kê của contig được lắp ráp đầu tiên

Mẫu	Số lượng gen	Số lượng transcript	GC (%)	N50 (bp)	Kích thước trung bình của contig (bp)	Tổng số base được lắp ráp (bp)
L4.1	70.526	106.276	40,59	1.127	728,62	77.434.527
L4.2	68.454	99.609	40,89	1.033	670,91	66.828.420
L4.3	81.480	121.554	40,08	1.325	785,70	95.505.500

C4.1	47.870	69.638	40,20	870	632,46	44.043.555
C4.2	63.268	100.651	39,01	1.023	694,26	69.878.264
C4.3	59.717	88.228	38,86	828	613,73	54.148.380

Bảng 3.5 Kết quả thống kê quá trình lắp ráp các đoạn đọc ở mẫu mô lá (L4.1) và thân rễ (C4.1)

Mẫu	Thông số lắp ráp	Tất cả các contig của transcript	Chỉ isoform dài nhất/"gen"
L4.1	Tổng số trinity "gen"	70.526	70.526
	Tổng số trinity transcript	106.276	70.526
	GC (%)	40,59	40,80
	N90	300	258
	N80	458	347
	N70	665	481
	N60	888	699
	N50	1.127	993
	N40	1.374	1.312
	N30	1.645	1.642
	N20	1.989	2.025
	N10	2.547	2.618
	Độ dài contig lớn nhất	7.794	7.794
	Độ dài contig nhỏ nhất	201	201
	Median của độ dài contig	460,0	356,0
	Độ dài contig trung bình	728,62	623,76
	Tổng số base được lắp ráp	77.434.527	43.990.999
C4.1	Tổng số trinity "gen"	47.870	47.870
	Tổng số trinity transcript	69.638	47.870
	GC (%)	40,20	40,55
	N90	291	263
	N80	430	361
	N70	587	512
	N60	736	697
	N50	870	861
	N40	1.008	1.015
	N30	1.147	1.161
	N20	1.315	1.336
	N10	1.568	1.590
	Độ dài contig lớn nhất	7.854	7.854

	Độ dài contig nhỏ nhất	201	201
	Median của độ dài contig	494,0	402,0
	Độ dài contig trung bình	632,46	589,77
	Tổng số base được lắp ráp	44.043.555	28.232.335

2.3 Kết quả phân nhóm các đoạn trình tự thành các unigene

Các đoạn đọc sau khi lắp ráp *de novo* được tiến hành phân nhóm để tìm ra các unigene sử dụng CD-HIT-EST. Đây là một thuật toán bắt đầu với trình tự đầu vào có kích thước lớn nhất được chọn làm nhóm đại diện đầu tiên và phân chia những trình tự còn lại thành trình tự đại diện hay dư thừa dựa trên độ tương đồng với các đại diện đang xét. Độ tương đồng trình tự được tính toán dựa trên số lượng các word chung bằng cách sử dụng word indexing và bảng đếm để lọc ra những so sánh trình tự không cần thiết và tính độ tương đồng. Các bản phiên mã có kích thước lớn nhất thuộc cùng một locus sau khi được phân nhóm sẽ được coi là các contig unigene. Bảng 3.6 thống kê các contig unigene sau khi phân nhóm bao gồm số lượng gen, số lượng bản phiên mã, tỉ lệ GC, chỉ số N50, kích thước trung bình của contig và tổng số base được lắp ráp. Các dữ liệu cho thấy, các mẫu mô khác nhau về số lượng gen (46.034-67.882), số lượng bản phiên mã (46.034-67.882) và số lượng base lắp ráp (27.641.662-50.440.507). Sự chênh lệch này chủ yếu do ảnh hưởng của số lượng đoạn đọc cũng như số lượng và kích thước các contig được lắp ráp.

Bảng 3.6 Kết quả thống kê của contig unigene

Mẫu	Số lượng gen	Số lượng transcript	GC (%)	N50 (bp)	Kích thước trung bình của contig (bp)	Tổng số base được lắp ráp (bp)
L4.1	67.882	67.882	40,73	1.022	635,07	43.109.582

L4.2	66.796	66.796	41,00	932	594,70	39.723.558
L4.3	77.381	77.381	39,97	1.120	651,85	50.440.507
C4.1	46.034	46.034	40,49	876	600,46	27.641.662
C4.2	59.818	59.818	38,90	961	615,62	36.825.282
C4.3	57.892	57.892	39,04	828	583,11	33.757.379

2.4 Kết quả dự đoán khung đọc mở

Bảng 3.7 thể hiện dữ liệu thống kê kết quả dự đoán ORF của các mẫu mô lá và thân rễ sâm Ngọc Linh 4 năm tuổi. Kết quả thống kê cho thấy, đối với 3 mẫu L4.1, L4.2 và L4.3, có khoảng 33,04-37,01% các unigene được dự đoán ORF. Trong số đó, có 92,65-96,17% được dự đoán có 1 ORF và 3,83-7,35% có nhiều ORF. Trong số 22.091-25.713 ORF được dự đoán có 32,68%-48,4% ORF hoàn chỉnh, 25,45-34,38% ORF thiếu bộ ba mở đầu, 5,76-7,72% ORF thiếu bộ ba kết thúc và 18,71-27,18% ORF thiếu cả 2 loại này. Đối với 3 mẫu C4.1, C4.2 và C4.3, có khoảng 23,22-36,18% các unigene được dự đoán ORF. Trong số đó, có 95,61-97,95% được dự đoán có 1 ORF và 2,05-4,39% có nhiều ORF. Trong số 14.513-20.695 ORF được dự đoán có 22,22-36,53% ORF hoàn chỉnh, 46,65- 65,9% ORF thiếu bộ ba mở đầu, 1,73-4,09% ORF thiếu bộ ba kết thúc và 10,15- 12,72% ORF thiếu cả 2 loại này. Kết quả, không có sự khác nhau nhiều giữa các mẫu dựa trên tỉ lệ tương đối giữa số lượng unigene và ORF.

Bảng 3.7 Kết quả thống kê dự đoán ORF của các mẫu mô lá và thân rễ sâm Ngọc Linh 4 năm tuổi

Mẫu	Số lượng unigen	Unigen dự đoán là ORF	Unigen dự đoán là đơn ORF	Unigen dự đoán là đa ORF	Số lượng ORF	Gen hoàn chỉnh	Một phần/ vùng mang mã	Một phần của đầu 5'	Một phần của đầu 3'
L4.1	67.882	22.426 (33,04%)	21.293 (94,95%)	1.133 (5,05%)	23.619	9.311 (39,42%)	5.659 (23,96%)	6.826 (28,9%)	1.823 (7,72%)
L4.2	66.796	24.722 (37,01%)	23.775 (96,17%)	947 (3,83%)	25.713	8.403 (32,68%)	6.989 (27,18%)	8.840 (34,38%)	1.481 (5,76%)
L4.3	77.381	20.458 (26,44%)	18.954 (92,65%)	1.504 (7,35%)	22.091	10.692 (48,4%)	4.134 (18,71%)	5.622 (25,45%)	1.643 (7,44%)
C4.1	46.034	16.653 (36,18%)	16.272 (97,71%)	381 (2,29%)	17.044	4.967 (29,14%)	1.939 (11,38%)	9.650 (56,62%)	488 (2,86%)
C4.2	59.818	13.887 (23,22%)	13.277 (95,61%)	610 (4,39%)	14.513	5.302 (36,53%)	1.846 (12,72%)	6.771 (46,65%)	594 (4,09%)
C4.3	57.892	20.270 (35,01%)	19,855 (97,95%)	415 (2,05%)	20.695	4.599 (22,22%)	2.101 (10,15%)	13.638 (65,9%)	357 (1,73%)

2.5 Kết quả ước lượng độ phong phú

Các đoạn đọc sau lọc chất lượng ở các mẫu mô lá và mô thân rễ sâm Ngọc Linh 4 năm tuổi được so sánh với trình tự tham chiếu của chính các mẫu này đã lắp ráp sử dụng phần mềm Bowtie. Để phân tích sự khác biệt biểu hiện gen, độ phong phú của các unigene giữa các mẫu được ước tính thông qua số lượng đoạn đọc sử dụng thuật toán RSEM. Bảng 3.8 thể hiện tỉ lệ lắp ráp các đoạn đọc của các mẫu mô lá và thân rễ sâm Ngọc Linh. Kết quả phân tích cho thấy, có 66,92-72,76% các đoạn đọc được lắp ráp vào hệ phiên mã tham chiếu của chính các mẫu này. Điều này cho thấy có 27,24-33,08% các đoạn đọc không được lắp ráp vào hệ gen tham chiếu.

Bảng 3.8 Tỉ lệ lắp ráp các đoạn đọc của mẫu mô lá và thân rễ sâm

Mẫu	Số lượng đoạn đọc được xử lý	Số lượng đoạn đọc được lắp ráp	Số lượng đoạn đọc không được lắp ráp
L4.1	89.141.174	59.649.098 (66,92%)	29.942.076 (33,08%)
L4.2	74.486.684	50.129.670 (67,30%)	24.357.014 (32,70%)
L4.3	65.523.252	45.102.590 (68,83%)	20.420.662 (31,17%)
C4.1	61.627.462	44.405.868 (72,06%)	17.221.594 (27,94%)
C4.2	76.350.630	52.520.512 (68,79%)	23.830.118 (31,21%)
C4.3	62.801.632	45.693.642 (72,76%)	17.107.990 (27,24%)

3. CHÚ GIẢI HỆ PHIÊN MÃ CỦA CÁC MÔ SÂM NGỌC LINH 4 NĂM TUỔI

3.1 Chú giải hệ phiên mã mô lá và thân rễ sâm Ngọc Linh 4 năm tuổi

Các unigene thu được ở các mẫu mô lá và thân rễ sâm Ngọc Linh 4 năm tuổi được chú giải sử dụng các cơ sở dữ liệu Gene Ontology (GO), Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (EggNOG), NCBI Nucleotide (NT), NCBI non-redundant Protein (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), Universal Protein Resource (UniProt) và Pfam.

3.2 Kết quả chú giải dựa trên cơ sở dữ liệu GO

Kết quả chú giải các transcript của hệ phiên mã mẫu mô lá và thân rễ sâm Ngọc Linh dựa trên sự tương đồng về mặt trình tự trên cơ sở dữ liệu GO sẽ tương ứng với các mã GO chứa thông tin về trình tự và chức năng của các gen hoặc các sản phẩm của gen tham chiếu. Các transcript đã chú giải được phân chia theo chức năng thành 3 nhóm chính: Chu trình sinh học (biological process - BP), thành phần tế bào (cellular

component - CC) và chức năng phân tử (molecular function - MF). Trong nghiên cứu này, kết quả chú giải chức năng cho tỉ lệ các transcript thuộc nhóm BP chiếm 13,80-18,09% ở mô lá cây 4 năm tuổi và 14,05-17,53% ở mô thân rễ cây 4 năm tuổi. Các unigene thuộc nhóm CC chiếm từ 15,15-19,37% ở mô lá cây 4 năm tuổi và 15,75-19,50% ở mô thân rễ cây 4 năm tuổi. Tỉ lệ này của các transcript thuộc nhóm MF là 13,41-17,35% ở mô lá cây 4 năm tuổi và 14,11-17,50% ở mô thân rễ cây 4 năm tuổi. Số lượng unigene không được chú giải chiếm từ 45,19-57,64% ở mô lá cây 4 năm tuổi và 45,46-56,08% ở mô thân rễ cây 4 năm tuổi. Có thể thấy, tỉ lệ các unigene không được chú giải tương đối cao. Điều này có thể giải thích do sự hạn chế về số lượng các gen/ sản phẩm của gen tham chiếu cho các loài thuộc chi *Panax* trên cơ sở dữ liệu. Hình 3.3 minh họa thông tin phân nhóm của các mã GO thu được sau quá trình chú giải mẫu mô L4.1 sâm 4 năm tuổi.

3.3 Kết quả chú giải dựa trên cơ sở dữ liệu EggNOG

Thông tin về tỉ lệ phân nhóm dựa theo chức năng của các unigene đã chú giải ở mẫu mô lá L4.1 dựa trên cơ sở dữ liệu EggNOG được minh họa trên Hình

3.4. Các unigene được chú giải trên cơ sở dữ liệu EggNOG được chia làm 3 nhóm chính: (1) Nhóm các gen có chức năng trong các quá trình sinh học và lưu trữ thông tin (information storage and processing); (2) Nhóm các gen liên quan đến chu trình và tín hiệu tế bào (cellular processes and signaling); (3) Nhóm các gen liên quan đến trao đổi chất (metabolism). Tuy nhiên, tỉ lệ các unigene không được chú giải vẫn tương đối cao do hạn chế về thông tin của các loài thuộc chi *Panax* trên cơ sở dữ liệu EggNOG.

3.4 Kết quả chú giải dựa trên cơ sở dữ liệu NT và NR của NCBI

Dữ liệu về trình tự nucleotide và trình tự amino acid trên cơ sở dữ liệu NT và NR của NCBI được sử dụng làm tham chiếu để chú giải các hệ phiên mã. Quá trình tìm kiếm tương đồng các unigene của hệ phiên mã sâm Ngọc Linh đưa ra kết quả là các mã NCBI đại diện tương ứng với các unigene so sánh, chứa thông tin về trình tự, tên chú giải, kích thước, độ bao phủ, tỉ lệ tương đồng, số lượng gap, độ tin cậy, vị trí khung đọc mở... (Bảng 3.).

3.5 Kết quả chú giải dựa trên cơ sở dữ liệu KEGG

Cơ sở dữ liệu KEGG chứa chủ yếu thông tin về trình tự, chức năng và các pathway liên quan của các protein tham chiếu. Bảng 3. trình bày kết quả chú giải hệ phiên mã của mẫu mô lá và thân rễ sâm Ngọc Linh đại diện. Hình 3.6 mô tả thông tin phân nhóm của các mã GO thu được sau quá trình chú giải các mẫu. Tương ứng với các unigene là các mã KEGG chứa thông tin về trình tự, tên chú giải, pathway và các thông tin so sánh khác như kích thước, độ bao phủ, tỉ lệ tương đồng, số lượng gap, độ tin cậy, vị trí khung đọc mở...

3.6 Kết quả chú giải dựa trên cơ sở dữ liệu UniProt

Dữ liệu về trình tự amino acid / protein trên cơ sở dữ liệu UniProt được sử dụng làm tham chiếu cho quá trình chú giải hệ phiên mã mẫu mô lá và thân rễ sâm Ngọc Linh. Bảng 3.1 trình bày kết quả chú giải hệ phiên mã đại diện. Tương ứng với các unigene là các mã UniProtKB chứa thông tin về trình tự và các thông tin so sánh khác như kích thước, độ bao phủ, tỉ lệ tương đồng, số lượng gap, độ tin cậy, vị trí khung đọc mở...

3.7 Kết quả chú giải dựa trên cơ sở dữ liệu Pfam

Cơ sở dữ liệu Pfam chứa thông tin về trình tự, chức năng và thông tin về các domain của các protein tham chiếu. Quá trình chú giải

các unigene của hệ phiên mã mẫu mô lá và thân rễ sâm Ngọc Linh đưa ra kết quả là các mã chú giải Pfam tương ứng với các unigene so sánh, chứa thông tin về trình tự, tên chú giải và các thông tin so sánh khác như kích thước, độ bao phủ, tỉ lệ tương đồng, số lượng gap, độ tin cậy, vị trí khung đọc mở... (Bảng 3.11).